DOCUMENT RESUME

ED 174 625 TM 008 176

AUTHOR Goodrich, Robert L.
TITLE Stochastic Models of Human Growth.
INSTITUTION Abt Associates, Inc. Cambridge, Mass.
PUB DATE Mar 78
NOTE 177p.

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Child Development; *Cognitive Measurement; Day Care Services; *Intellectual Development; Intelligence Quotient; Intelligence Tests; *Longitudinal Studies; *Mathematical Models; Predictive Measurement; *Research Methodology; Research Reviews (Publications); Statistical Studies
IDENTIFIERS *Fels Longitudinal Programs; *National Day Care Study

ABSTRACT

Stochastic difference equations of the Box-Jenkins form provide an adequate family of models on which to base the stochastic theory of human growth processes, but conventional time series identification methods do not apply to available data sets. A method to identify structure and parameters of stochastic difference equation models of human growth is presented. The practicality of the method is demonstrated in two applications to human growth data. The data sets are: (1) IQ data for 80 subjects over a period of 14 years, from the Fels Institute Longitudinal Sample. (2) Preschool readiness test data for 110 subjects over a period of two years, from the National Day Care Center (NDCS) Longitudinal Sample. Applicability of stochastic system theory to cognitive growth problems is established. In the Fels data, the existence of a stochastic component to growth is demonstrated. The findings support the psychological hypothesis that IQ is, not as much a stable human trait as it is a dynamic phenomenon that exhibits real, persistent changes in response to changing environments. An upper bound to "heritability" of IQ in the Fels population was computed; it is substantially lower than other figures quoted in the literature. (Author/RD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert L. Woodruff

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Abt Associates Inc., Cambridge, Massachusetts

TM008 176

ABT ASSOCIATES INC
55 WHEELER STREET, CAMBRIDGE, MASSACHUSETTS 02138
TELEPHONE • AREA 617-492-7100

Stochastic Models of Human Growth

March, 1978

Robert L. Goodrich

Section 27.12

Prepared for

The American Educational Research Association
1978 Annual Meeting
Toronto, Ontario

Copyright Reserved by the Author

This is an early draft of a
thesis to be presented to

Department of Applied Sciences
Harvard University

Printed in U.S.A.

Preface

This thesis was motivated by problems encountered in my work as research director of the National Day Care Study (NDCS), a federally funded study of day care center effects on child development. I began to think of repeated longitudinal tests administered to children as short sequences of imperfect measures of cognitive growth trajectories. It was only a step beyond this view to that of considering that the trajectories were generated as realizations of nonstationary stochastic processes. That is the main substance of this thesis. I would like to acknowledge my debt to Tony Bryk and Herbert Weisberg, who aroused my interest in growth curve dynamics, and to all those who helped to provide the environment at Abt Associates Inc. that nurtured these ideas, especially Dr. Richard Ruopp, who created our research team and was the chief architect of the NDCS. I would also like to thank Mr. Allen Smith, our government project officer and initiator of the study.

While the NDCS provided the substantive motivation for this study, it was my teachers at Harvard who helped me to see it in quantitative terms. My thesis director, Professor Peter E. Caines, has left many marks on this work as the result of technical suggestions that worked. This thesis is the product of many hours of conversation with Peter Caines. Each conversation resulted in new insights and new enthusiasm for the project. I would especially like to thank him for taking on the project in the first place: day care educational research cannot have sounded very exciting to him in the original formulation of this thesis.

I would also like to thank Professor Y.C. Ho for his encouragement and support for a student who must have seemed a little slow: this thesis was submitted twenty years after my admission to the doctoral program in the (then) Division of Engineering and Applied Physics.

I would also like to thank Professor Richard Light for encouragement and useful technical conversations, and Professor Myron Fiering for important suggestions about the presentation and statistical issues.

In preparing the typewritten copy my debt is to Diane Almeida, who must have grown weary with repeated revisions. I would like to acknowledge the competent support and cooperation of the computer operators at Smithsonian Astrophysical Observatory, Cambridge, Mass.

Finally, I owe a debt to Drs. Robert McCall and Alex Roche of the Fels Institute, Yellow Springs, Ohio. Dr. McCall evoked my interest in the Fels data through his important SRCD monograph, Developmental Changes in Mental Performance. He and Dr. Roche were kind enough to allow me access to the original data, as representatives of Fels Institute.

I also appreciate the permission of the Society for Research and Child Development (SRCD) to use a figure from Dr. McCall's monograph, reproduced as Figure (4.1.1) in this thesis.

Table of Contents

		Page
Synopsis		v
Chapter I	Introduction	
	1.1 Summary of Methodology	1-1
	1.2 Relation to Previous Research	1-4
	1.3 Data Analysis	1-8
Chapter II	Definition of Model	
	2.1 Introduction	2-1
	2.2 Model of Input Process and Growth Dynamics	2-2
	2.3 Measurement Error Model	2-11
	2.4 State Space Representation	2-15
	2.5 Formal Extension of SS Model to Include Individual Random Coefficients	2-16
	2.6 Probability Theory Foundation	2-17
Chapter III	Model Identification	
	3.1 Introduction	3-1
	3.2 Maximum Likelihood Estimates of Parameters	3-1
	3.3 Consistency of Global ML Estimates	3-7
	3.4 Consistency of Stepwise ML Estimates	3-14
	3.5 Model Identifiability (T, Ω)	3-19
	3.6 Local Identifiability and the Hessian Matrix	3-21
	3.7 Structure Determination	3-23
	3.8 Implementation of Methods	3-25
	3.9 Prediction Error Formulation	3-27
Chapter IV	Analysis of Fels Institute Longitudinal Data	
	4.1 Introduction	4-1
	4.2 Conclusions of McCall et al. and Previous Researchers	4-4
	4.3 Hypotheses of this Study	4-6
	4.4 Identification of System	4-8
	4.5 Model Validation	4-15
	4.6 Substantive Conclusions	4-22

Table of Contents (cont'd.)

		Page
Chapter V	Analysis of National Day Care Study Longitudinal Data	
	5.1 Introduction	5-1
	5.2 Identification of Fluctuation Model	5-3
	5.3 Estimation of Mean Growth Curve	5-12
	5.4 Construction of Generalized Change Scores	5-14
Chapter VI	Summary and General Conclusions	
	6.1 Summary of Findings	6-1
	6.2 Useful Extensions	6-2
	6.3 Other Applications	6-4
Appendices		
Appendix 1	Calculations of Gradient and Hessian Matrix	A1-1
Appendix 2	Calculation of Filter Gain Gradient	A2-1
Appendix 3	An Analytic Example: Pre/post Treatment Design	A3-1
Appendix 4	Proof of Main Identifiability Theorem	A4-1
Appendix 5	Identifiability of Random Trend Parameters σ^2_R and D_R	A5-1
Appendix 6	Estimation of Mean Growth Curve	A6-1
Appendix 7	Local Identifiability and the Hessian Matrix	A7-1
Appendix 8	Listing of Computer Program GCURVE	A8-1
Bibliography		B-1

Synopsis

This thesis was motivated by the need for a dynamic stochastic theory of human growth processes. Stochastic difference equations of the Box-Jenkins form (ARMA models) provide an adequate family of models on which to base the theory, but conventional time series identification methods do not apply to available data sets. Not only are growth data nonstationary, remaining dependent on initial conditions, but they usually involve short measurement sequences (although the cross-sectional sample size may be large).

This thesis presents a method to identify structure and parameters of stochastic difference equation models of human growth. The methods also apply to cross-sectional data sets that occur frequently in econometrics, sociology, and ecology. The practicality of the method is demonstrated in two applications to human growth data. The data sets are: (1) IQ data for 80 subjects over a period of 14 years, from the Fels Institute Longitudinal Sample. (2) Preschool readiness test data for 110 subjects over a period of two years, from the National Day Care Study (NDCS) Longitudinal Sample. The Fels data have not been previously analyzed from this point of view, although they have been studied by many authors, for instance McCall et al. [7] and Sontog et al. [44]. The NDCS data have not been previously analyzed.

We present a maximum likelihood method to identify model parameters in a state space formulation. The likelihood function and its derivatives are calculated by using the Kalman filter technique. Applications of the Kalman filter identification method to conventional time series data are now standard in the literature, but application to cross-sectional problems is apparently new. A FORTRAN version of the technique was written, debugged, tested and applied to the data sets cited.

We demonstrate~~d~~ that estimators obtained by this method are consistent, provided the model is identifiable, i.e., the parameters that define given second order statistics of growth sequences are unique. We present new theorems that state necessary and sufficient conditions for identifiability. Because the identification procedure involves the Ricatti equation, we have termed the associated problem the inverse Ricatti problem.

Local identifiability is also investigated. In particular the work of Tse [70] and F. Fisher [60] is extended in relating (1) nonsingularity of the information matrix (Hessian) to (2) identifiability of the process. The theorem stated presents a qualifying condition under which these two conditions are equivalent. The theorem is of practical value for model order determination from sample data.

This thesis presents new results in psychology. Applicability of stochastic systems theory to cognitive growth problems is established. In the Fels data the existence of a stochastic component to growth is demonstrated. Our findings support the psychological hypothesis that IQ is less a stable human trait than a dynamic phenomenon that exhibits real, persistent changes in response to changing environments. We compute an upper bound to "heritability" of IQ in the Fels population that is substantially smaller than others quoted in the literature (for a variety of populations).

The methodology is also applied to the very short (4 points) time series obtained in the National Day Care Study. The results of the identification of these data are used to construct a measure of change (Generalized Change Score) of a developmental variable that is useful for investigation of effects of day care environments on cognitive growth. The results of this evaluation are not, however, presented in this thesis.

Chapter I Introduction

1.1 Summary of Methodology

This thesis explores the use of dynamic concepts from systems theory for the estimation of human growth curves, and of treatment effects when subjects are growing. Nearly all previous work has taken an essentially static approach to this dynamic problem. The present work concentrates on the dynamics of the development process de ite obvious difficulties in identifying structure and parameters from the sparse longitudinal data that are usually available.

The fundamental concept of the paper is that of the developmental trajectory, which expresses the growth of variables such as mental age, height, or weight. The growth process is subject to the forces of maturation, chance events in the environment, specific nurturing, and intervention. This picture of the process leads to the basic hypothesis of the study, that human growth curves may be usefully described as stochastic difference equations involving three distinct sources of randomness: variance due to fluctuating environmental conditions, variance in individual abilities or environmental patterns, and random measurement error. Properties to be studied are both systematic and random, cross-sectional and longitudinal. Thus an ideal data set consists of a large number of individual time series forming a sample of developmental trajectories.

The theoretical framework for the distribution of trajectories is based on state space models that extend the autoregressive moving average (ARMA) model discussed for instance by Box and Jenkins [3]. Justification for this choice is provided in Chapter II. However, the identification methods of Box and Jenkins are not directly applicable,

since they deal with time series of essentially infinite length, assuming stationarity. What has been lacking is a method to analyze an entire cross-section of short, non-stationary time series simultaneously, considering them as cross-sectional rather than longitudinal samples. The major contribution of this paper is a statistically valid method for identification and estimation of such cross-sectional/longitudinal systems.

The method, which uses the principle of the Kalman filter, was programmed in FORTRAN for the CDC 6400 computer and was used in the analysis of two data sets. They are: the Fels Institute longitudinal IQ data (80 subjects, 14 years, 17 measurements) that formed the basis for a monograph by McCall, Appelbaum, and Hogarty [7]; and the National Day Care Study preschool cognitive development data (250 subjects, 2 years, 4 measurements).

The technical approach of this thesis may be summarized as follows:

(1) Each subject of a cross-sectional sample is assumed to be experiencing a developmental process containing both deterministic and stochastic components, manifested as a longitudinal sequence of univariate measurements. Each such sequence represents a developmental trajectory.

(2) The growth model to be constructed depends on the concepts of <u>state</u> and <u>state-space</u>. In general terms the state of each subject represents the sum total of genotypic and phenotypic traits that influence the trajectory at a particular point in time. The true state is <u>unobservable</u> and inferences about it must be made through the intermediary of inexact measures $\{z_k\}$. The state evolves dynamically as a function of maturation processes, environmental encounters and similar mechanisms.

(The notion of state as we use it is to be distinguished from the psychological construct of the same name. In psychology the notion of state refers rather to the immediate arousal condition of the organism, as for instance the heart rate level as a function of autonomic nervous system control and activity level, or of attention in response to environmental stimulation. Furthermore psychologists often identify specific measures with state, e.g., eye gaze direction with attention, while we will always carefully distinguish measures from unobservable internal processes.)

It will be assumed that subjects have been undergoing the growth process for some unspecified time so that the state at the time of first measurement includes accumulated effects of previous growth. Therefore an important part of the model to be identified is the distribution of initial states across the sample and by inference in the underlying population. The initial state contains all essential information available about future growth. If there are available additional data that are predictive of future growth then these data must also be included in the initial state.

(3) The development of the state is considered to be driven by three simultaneous input processes. These are (1) a stochastic input W representing random environmental events and/or random individual responses, (2) a deterministic input U representing stable characteristics of the environment extending over all subjects, and (3) a random trend input R representing environmental/ genotypic influences that are constant for each subject but are randomly distributed across subjects.

(4) The family of dynamic models describing evolution of the growth process consists of state space extensions of stationary and nonstationary stochastic

difference equations from the ARMA (Autoregressive Moving Average) family, driven by the three input processes discussed in (3). These models encompass a large group of growth processes.

(5) Structure determination and parameter identification depend upon the use of Kalman filter techniques, which are used to calculate the exact likelihood function and its derivatives. The estimation technique employed requires the rather severe assumption of normality for all input processes and initial conditions as a strict maximum likelihood method. As the cross-sectional sample size grows large, maximum likelihood parameter estimates converge to true values provided that certain identifiability conditions are satisfied.

(6) Identifiability is defined in terms of uniqueness of parameters that generate theoretical first and second moments of measurements. The conditions under which identifiability holds are investigated in detail. These conditions include minimum measurement sequence lengths and certain algebraic requirements. Several new theorems are presented and proved, extending previously known results in stochastic systems identification theory.

The technical development is presented in Chapters II and III, supported by Appendices 1-7. Chapters IV and V provide analyses of the Fels and NDCS data. The computer program is provided in Appendix 8.

1.2 Relation to Previous Research

The methodology of this thesis relates to previous work in several areas, including system identification theory, "growth system theory,"* the nonequivalent control

*"Growth System theory" is not a generally used term. We have coined the phrase to indicate the work begun by Bryk and Weisberg (described below) and extended in this thesis.

group problem, and random coefficient regression. These connections will be discussed in turn.

System Identification Theory

This work builds on the main stream of system identification theory developed by many authors in the past two decades. Of particular relevance are Akaike [56,65], Anderson [78], Caines [12,35,77], Faurre [34,80,82], Ho and Kalman [64], Kalman [29,30,79], Mehra [26,27,28,66], Rissanen [35,45,63], Tse [70,71], and Youla [75]. This thesis extends the conventional longitudinal formulation of the problem to the nonstationary, cross-sectional case that we have described. The principal new contribution to systems theory consists of the series of theorems regarding conditions for identifiability of stochastic systems from repeated short sequences of measurements.

Bryk and Weisberg's "Growth System Theory"

The motivation for this research came from the need in the National Day Care Study to find methods of evaluating environmental effects of day care on cognitive development of preschool children. While the NDCS evaluation raises many methodological problems, this thesis focuses on just one of them: the essentially dynamic (and stochastic) nature of the growth process under consideration, a nature not well accomodated by existing static analytic strategies. Our interest in this particular aspect of the problem was awakened by two papers published by Bryk and Weisberg [1,2], in which two fundamental ideas were set forth. The first of these is the conceptualization of the developmental trajectory as one sample from a growth system for the population under consideration. The second is the concept of "value added" as a measure of treatment effect in growth systems. Value-

added is, in heuristic terms, the increment of growth attributable to educational intervention. Bryk and Weisberg's research was among the first to use the developmental trajectory as the primary object of analysis.

An introduction to the growth system concepts considered in this thesis is provided in Anthony Bryk's 1977 thesis, An Investigation of the Effectiveness of Alternative Statistical Adjustment Strategies in the Analysis of Quasi-Experimental Growth Data [52]. Methodologically the present thesis can be regarded as an extension and generalization of Bryk's work in the following directions:

(1) The class of growth models considered is enlarged to include the large family of models that can be generated by linear difference equations of finite order, and time varying input. This family includes finite difference versions of Bryk's linear and negative exponential models as special cases. Bryk's models deal with variables changing in continuous time.

(2) The models considered in this thesis are stochastic, in contrast to Bryk's deterministic models. Bryk remarks that his choice of deterministic model "should not be construed as a statement about the ultimate nature of growth phenomena," and goes on the say, "While we need to recognize in our models the stochastic nature of the measurements at each point this is not the essential feature which I wish to explore. Rather, I am interested primarily in modeling whatever trend exists within each individual's development over time. It seems reasonable then to conceive of smooth individual growth curves which weave their way through the temporal idiosyncracies which characterize real growth. To the extent that the concept of a systematic component to individual growth over time is reasonable, then a basically deterministic position is conceptually adequate."

Inclusion of stochastic effects gives us the opportunity to test Bryk's hypothesis about the ultimate irrelevance of "temporal idiosyncracies."

(3) Bryk's concept of "value-added" as a measure of treatment effect is generalized, leading to definition of a generalized change score (CGS) for use in treatment evaluation. Bryk and Weisberg [2] demonstrated the dependency of change analysis on identification of the underlying growth system, but did not explore the problem of system identification or deal with stochastic effects.

This work is distinct from growth curve theory as usually conducted in biometrics. These techniques, are represented (for example) by the curve fitting methods of Bock and his colleagues [39,40,41,42], where sums of logistic curves are fitted by weighted least squares. Such methods provide a useful parameterization of individual growth curves but they are neither stochastic nor dynamic and are thus not related to the present theory.

Random Coefficient Regression

The theory to be presented is based on the hypothesis that growth trajectories are represented by stochastic difference equations with subject-to-subject variations that are partially random. Subject-to-subject variations are modeled by randomization of system parameters across individuals. Such models have been much used lately in econometric theory, and are known as RCR (Random Coeffcient Regression) models. A large literature on the estimation of these models is available (Rosenberg, [17]; Zellner, [18]; Swamy, [19]). While the present work does not build upon this literature directly, it is substantively related.

1.3 Data Analysis

The methodology developed in this thesis applies most directly to data sets with the following characteristics. First, the data are univariate measurements, error corrupted, of human developmental characteristic such as IQ or stature or of similar non-stationary processes. Second, the data form time series of measurements, not necessarily at equal intervals. Third, there must be available a relatively large sample of such short time series, preferably at least 50. The Fels longitudinal data (80 subjects, 17 measurements over 14 years), to be discussed, are ideal for our purposes, while the National Day Care Study data (110 subjects, 4 measurements over 2 years) are rather short, but still analyzable.

The Fels data ~~are a portion of~~ consist of sequences of Stanford-Binet IQ measurements the Fels Institute Longitudinal Sample, studied by many researchers [7,44]. In analyzing them we will be interested in statistical distributions of "normal" development of intelligence, rather than in treatment effects. One of the most important and relevant previous analyses was conducted by McCall, Appelbaum, and Hogarty in their monograph Developmental Changes in Mental Performance [7]. Their interest was primarily in discerning systematic patterns of growth and the relation of these patterns to environmental characteristics. They regard short term fluctuations about longitudinal trends as nuisance variables that tend to obscure long term growth patterns. Our analysis is different in that these fluctuations are of primary interest in identification of the growth model. While our orientation is different, this work supports in a general way the basic hypothesis of McCall et al., who believe that patterns of growth are associated with different patterns of environmental conditions.

The second data set is from the National Day Care Study (NDCS), a federally funded study of about 2000 three and four year olds in 64 day care centers in Atlanta,

Detroit, and Seattle. The data set analyzed in this thesis consists of Preschool Inventory Test* (PSI) scores for 110 children, measured at four times over the course of two years. The NDCS data illustrate the advantages of obtaining more than the two points of longitudinal data usual in educational evaluation studies. We will concentrate on statistics of growth trajectories and the problem of measurement of change, rather then on day care effects on children. The analysis leads to construction of a generalized change score (GCS) which has proved useful in the NDCS analysis (not reported here).

*The PSI is a set of 64 cognitively oriented items developed by Bettye Caldwell [49] for the Headstart evaluation. The longitudinal NDCS data involves a shorter (32 items) version of the PSI originating in the Follow Through Evaluation.

Chapter II Definition of Model

2.1 Introduction

In this chapter we present the stochastic model of growth to be used throughout this thesis. The model is based on an extensive literature in engineering and mathematical statistics, centering around the topics of autoregressive moving average (ARMA) processes, and state space theory. Neither topic has been applied extensively in psychology, although AR (autoregressive) processes are currently recieving some attention, for instance from Campbell [53]. The reader who is unfamiliar with systems theory models is referred to Box and Jenkins [3] for background material in ARMA processes and Åström [13] for more general state space theory.

Our growth process model includes the following specific components, linked in a state space representation:

(1) A model of the genotypic/environmental input processes that "drive" development.

(2) A state space model of growth process dynamics based on the ARMA model group discussed by Box and Jenkins [3], and extended to include both stochastic and deterministic inputs.

(3) A linear model of the measurement process relating "true" to observable variables, based on classical test theory as presented by Lord and Novick [33].

The model will be developed first as an ARMA model of the "true" process and will subsequently be extended to a state space formulation. Figure 1 provides a diagram of the model to be presented, in ARMA representation.

2.2 Model of Input Process and Growth Dynamics

Input Process

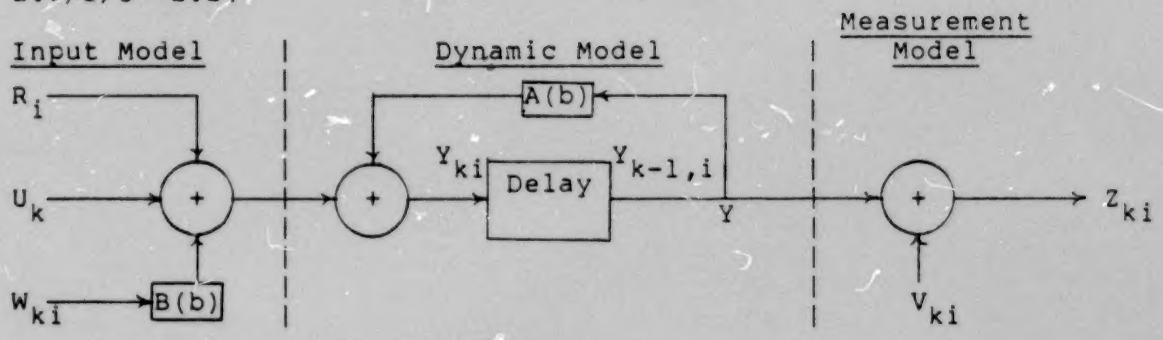
The growth model is driven by three forms of genotypic/environmental input with different patterns of statistical correlations. Figure 2 presents examples of these processes across three subjects and seven points in time. These inputs are defined as follows:

(1) W_{ki} is the stochastic input, uncorrelated across time periods (k) and across individuals (i), representing random fluctuations in environmental input and developmental response.

(2) U_k is the deterministic input, representing stable features of the environment and of intentional environmental inputs. In Figure 1, U_k is constant for k=1,2,3 and rises to a higher constant value for k=4,5,6,7, representing an intervention occurring between times 3 and 4. The sequence $\{U_k\}$ are unknown parameters to be estimated.

(3) R_i is the trend input constant for each individual but randomly distributed across individuals. R_i represents nurturing conditions and genotypic influences peculiar to individuals. The presence of this effect is associated with individual-to-individual variations in growth trends.

Figure 3 presents an example of measured growth trajectories that might be driven by such input processes. The fluctuation of each trajectory about its own trend is due to the stochastic input W and to measurement error, while the overall upward trend is driven by the deterministic input U. Variations of individual trends about the



R_i = Trend input specific to case i.

U_k = Deterministic input at time k.

W_{ki} = Stochastic input at time k, case i.

B(b) = Moving Average operator, converting uncorrelated input W to correlated input B(b)W.

A(b) = Autoregressive operator, representing system response.

Y_{ki} = True output at time k, case i.

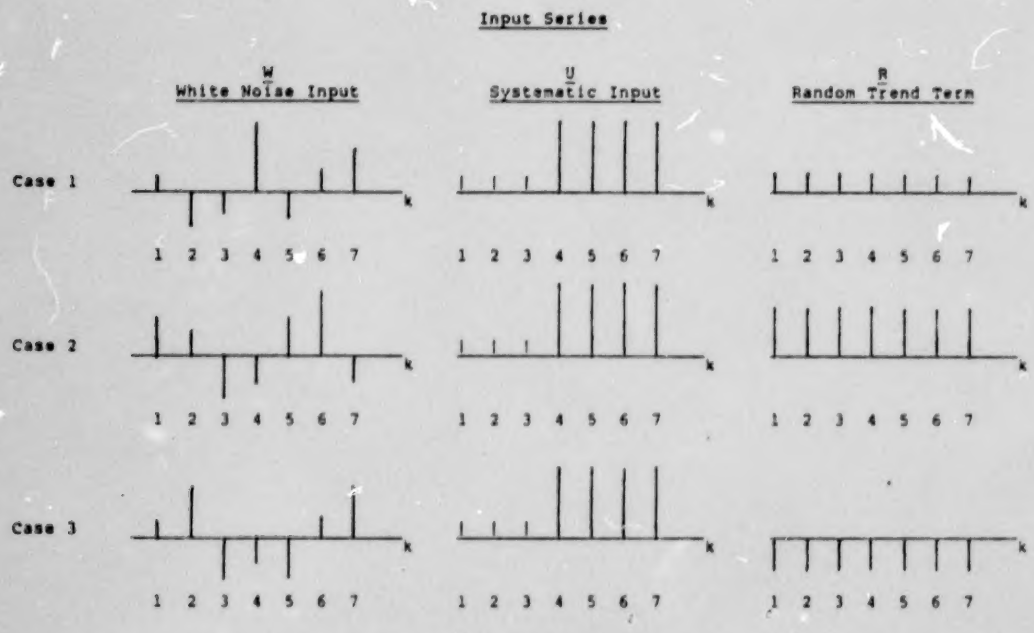
V_{ki} = Measurement error at time k, case i.

Z_{ki} = Measured output at time k, case i.

Block Diagram of Growth Model

Figure 1

cross-sectional trend are caused by variations in the trend input R, illustrating the divergence effect sometimes termed "fan spread".



Input Series for Three Sample Cases Illustrating Patterns of Variation

Figure 2

Growth Dynamics

Although the development process may be very complex measurements are assumed unidimensional. For instance cognitive development may be assessed with the Stanford-Binet test, a unidimensional measure. We will represent the true value of the developmental variable as a sequence $\{Y_k\}$, where the case index i has been suppressed for simplicity.

The general ARMA (p,q) process driven by these inputs is

$$Y_{k+1} = A_1 Y_k + A_2 Y_{k-1} + \ldots + A_p Y_{k+1-p} + W_k + B_1 W_{k-1} + \ldots + B_q W_{k-q} + U_k + R . \quad (1)$$

In a convenient symbolic form, (1) can be rewritten as

$$\boxed{\text{ARMA:} \quad A(b) Y_{k+1} = B(b) W_k + U_k + R} \quad (2)$$

where A(b) and B(b) represent operator polynomials in b, the backwards time operator:

$$\text{AR terms:} \quad A(b) = 1 - (A_1 b + A_2 b^2 + \ldots + A_p b^p) \quad (3)$$

$$\text{MA terms:} \quad B(b) = 1 + B_1 b + \ldots B_q b^q . \quad (4)$$

The following technical comments support our choice of the model given by Equation (2).

(1) ARMA/State Space Foundation

We cite important equivalences among broad groups of models to be considered. As discussed by Akaike [32], Faurre [34], and others, the following are equivalent families:

(I) Stationary time series with finite order state space representations.

(II) Stationary ARMA processes.

(III) Stationary series with rational spectra.

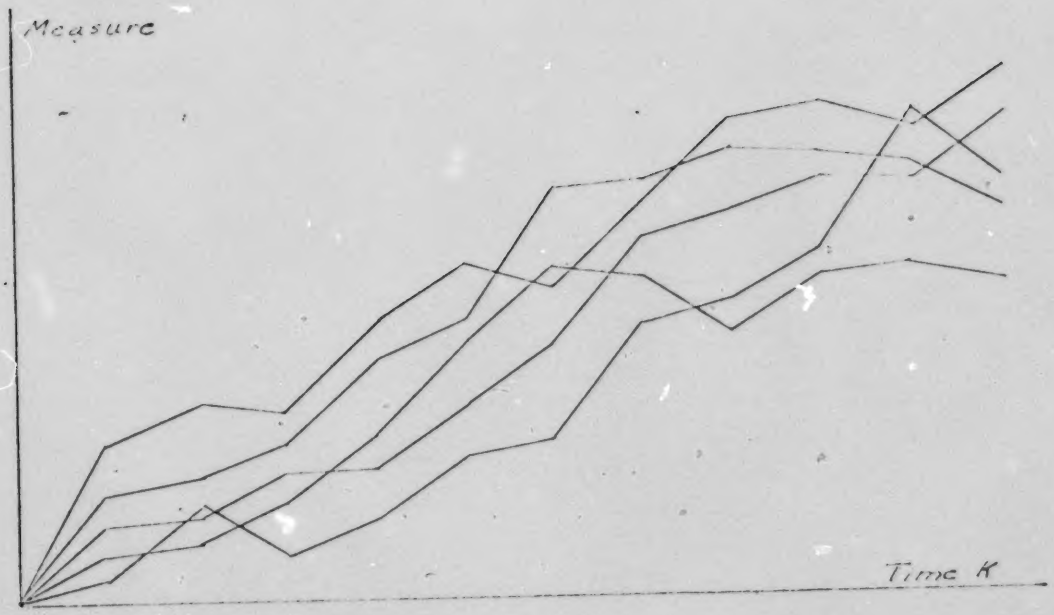


Figure 3. Sample of Measured Growth Trajectories (Fictitious Example)

The models we use are asymptotically equal to the types of processes just cited but also include transient non-stationary components reflecting system growth. We could describe more general stochastic processes by allowing time variation of parameters, but resulting models would be virtually unidentifiable. Practically speaking, we can do no better in modeling stochastic aspects of processes to be discussed.

(2) Inclusion of MA terms

It is common in economics and social science to consider the pure autoregressive AR(p) system obtained by setting $B_1, B_2, .., B_q$ to zero, partly because such systems can be estimated by ordinary regression. However, we include MA(q) terms as well for the following reasons.

(1) The use of MA terms may yield model identifications with many fewer parameters. For instance the MA(1) process is technically equivalent to an AR(∞) process and practically speaking might be approximated by an AR(10) process. Simpler models offer more hope for meaningful substantive inferences about the process.

(2) Unlike the case in most applications, we wish to represent mean growth and stochastic fluctuations around the mean. Retention of both MA and AR terms allows much more flexibility in fitting both these aspects simultaneously, assuming simple deterministic input sequences $\{U_k\}$.

(3) Form of Deterministic Input U

The unknown deterministic input sequence $\{U_k\}$ has a particularly simple form in equation (1). It might alternatively have been specified as a lagged sum of terms such as those for W_k in the same equation. This complicated specification is not mathematically distinct from the simpler version however, since each sum

$$U_k + C_1 U_{k-1} + C_2 U_{k-2} + .. C_s U_{k+1-s} \qquad (5A)$$

may be transformed by the equation

$$U'_k = U_k + C_1 U_{k-1} + .. + C_s U_{k+1-s} \qquad (5B)$$

to the simpler representation, fully equivalent to (5A). There is little motivation to introduce such additional complexity since it does not expand the class of describable growth processes.

(4) Stationarity and Stability

Box and Jenkins discuss unstable ARIMA (p,d,q) models involving d-fold differencing of the original sequence $\{Y_k\}$. Our state space identification techniques make this generic distinction awkward. Therefore we do not distinguish ARIMA generically from asymptotically unstable ARMA models. ARIMA models are seen as ARMA models for which one or more roots of A(b) lie on the unit circle, i.e., as specific subsets of points in a parameter space including ARMA, ARIMA and explosively unstable models. However we will sometimes use the term ARIMA for its

descriptive utility. To distinguish an ARIMA from a ARMA model we must set up a statistical hypothesis test involving parameter values.*

Note that by their very nature, growth processes cannot be stationary. An exception is the Fels data set involving IQ scores. Since IQ scores are normed at 100 points for each age group in the general population, the general population mean is fixed over time, although the mean for a specific population might exhibit growth or decline.

We will not consider models for which roots of A(b) lie outside the unit circle, in order to exclude explosively unstable models such as geometrical growth processes. For growth processes to be considered this restriction poses no problems. The limitation has the advantage that each process admissible under the restriction (and that of invertibility, to be discussed later) possesses a unique asymptotic spectrum, so that models are identifiable even in steady state.

(5) Divergence of Stochastic Growth Trajectories

Human growth trajectories are in general observed to drift apart deterministically and/or stochastically, in much the same pattern exhibited in Figure 2. This phenomenon is called "fan spread" by Bryk and Weisberg, although the original use of the term by Campbell [16] was more restricted. In our model as postulated fan spread is modeled in two forms:

(1) Individual to individual differences in environmental/gentoypic input level are described by variations in the input term R,

*See Chapter IV, where such a test is carried out for the Fels longitudinal data.

as discussed earlier.* The effect of R, assumed randomly distributed across the sample, is to drive the sample into individually divergent growth patterns. The estimation problem focuses on estimating the variance of R and its covariance with other parts of the initial state. A calculational by-product is the set of estimated levels of R, individual by individual, but these estimates are not consistent in cross-sectional sample size.

(2) Even in the absence of a random trend term R, effects of the random environmental input W accumulate to produce growing cross-sectional variance in trajectories. If the basic ARMA model is asymptotically stable the variance approaches a limiting value, but if the model is unstable individuals drift apart indefinitely in a random walk process. The resulting sample of trajectories may be difficult to distinguish from a sample from a stable system with random trend term R. (An example of this difficulty is presented in Chapter IV).

---

*Although the effect is random from individual to individual, it belongs to the deterministic part of the Wold decomposition [37]. Stochastic variation of parameters among individuals is interesting theoretically, and important to evaluation methodologists. The effect is to produce widening bundles of developmental curves even when all subjects are randomly selected from homogeneous demographic groups. These effects are one source of the well known problems of aggregation bias. Until recently no theory has existed for estimation of such systems, but the same problems were recognized in econometrics and led to development of RCR (Random Coefficient Regression) models. This methodology (Rosenberg, [17]; Zellner, [18]; Swamy, [19]) was not used in our development, however.

(6) Deterministic Growth Divergence

Except in the most limited way, (development of Generalized Change Score) to be described in Chapter 5, we do not consider deterministic differences in growth across subgroups of the sample. All subjects are assumed to experience the same input sequence $\{u_k\}$. In limiting the model in this way we exclude the possibility of a systematic analysis of interventional experiments with distinct treatment and control groups. Although we do not extend the model to include such interesting problems, the directions of such an extension should be fairly clear.

(7) Age Distribution of Subjects

There are potentially two time scales in the data sets to be considered in this thesis corresponding to (1) order within the measurement sequence $\{z_k\}$ and (2) age of subject at first measurement. For stationary systems, the age scale can be ignored. We will make the explicit assumption that all subjects are the same age, so that each measurement sequence refers to the same age span of the non-stationary development process. An extension to the case of a range of ages is not conceptually difficult but is not carried out in this thesis.

2.3 Measurement Error Model

The model (2.2.2) is assumed to hold for a developmental variable observable only in the presence of measurement error. We accept the assumptions of the classical theory of tests as discussed, for instance, by Lord and Novick ([33], Chapters 2 and 3). In this theory the true score is defined as the expected value of the observed score

~~score is defined as the expected value of the observed score~~ [33] Def.2.3.1). This definition avoids the necessity to explain the meaning of the measurement in any deeper way: IQ, for instance, is defined as the variable that IQ tests measure. This definition also establishes the scaling of the true score. The assumptions of the classical theory, on the other hand, require commitments to certain specific beliefs about the measurement process.

By definition of the true score Y_k,

$$Z_k = Y_k + V_k,$$

$$E(Z_k) = Y_k$$

$$E(V_k) = 0$$

where Z_k is the measurement and V_k the error.

By assumption,

$$E\ V_k V_m = 0 \qquad \text{if } k \neq m$$

$$E\ Y_k V_m = 0$$

$$E\ V_k^2 < \infty$$

$$E\ Y_k^2 < \infty$$

Distributional assumptions are not required in the classical theory, but in constructing the likelihood function we assume that Y_k and V_k are normally distributed.

Equivalence to Error-Free ARMA Model

The measurement model and the true score model form the following coupled system:

$$\boxed{\begin{aligned} A(b)Y_{k+1} &= B(b)W_k + U_k + R \\ Z_k &= Y_k + V_k \end{aligned}} \qquad (1)$$

Equations (1) and (2) can be combined to obtain

$$A(b)Z_{k+1} = \underbrace{B(b)W_k + U_k + R}_{\text{Order } q} + \underbrace{A(b)V_k}_{\text{Order } p} \qquad (2)$$

As discussed by Box and Jenkins ([3], pp.121-125), process (1) can be converted to an ARMA (p,r) process with r = Max(p,q). This may be seen by noticing that the RHS process $B(b)W_k + A(b)V_k$ is uncorrelated for lags exceeding r = Max(p,q), and is thus representable as MA(r). Thus:

$$\boxed{\begin{aligned} Y_k &\sim \text{ARMA}(p,q) \\ Z_k &= Y_k + V_k \end{aligned}} \equiv \boxed{Z_k \sim \text{ARMA}(p,\text{Max}(p,q))} \qquad (3)$$

It is easy to calculate new ARMA parameters in terms of the old by equating stationary spectra of the processes.

Uniqueness of Model

A spectral technique can be used to establish the following lemma regarding uniqueness of the model (1), which we assume to be stable and thus to posses an asymptotic stationary spectrum.

Lemma 1 Let p>q. Then model (1) parameters ($A, B, \sigma^2_w, \sigma^2_v$) are determined uniquely by the stationary spectrum, provided we require that the roots of A(b) lie inside te unit circle and those of B(b) lie on the unit disk.

Let $p \leq q$. Then A(b) is determined uniquely but B(b), σ^2_w, σ^2_v are not locally unique.

Proof: The stationary spectrum for the model (1) can be written

$$\Lambda(b) = \frac{B(b)B(b^{-1})\,\sigma^2_w + A(b)A(b^{-1})\,\sigma^2_v}{A(b)A(b^{-1})} \tag{4}$$

Suppose $\Lambda(b)$ is given as a rational function with no pole-zero cancellations. Then A(b) can always be determined uniquely if we require that its zeroes lie inside the unit circle, since the denominator of $\Lambda(b)$ is assumed known.

Now let $p>q$, so that the term of highest power of b in the numerator is $A_p\sigma^2_v b^p$. Since A_p is determined uniquely from the denominator, σ^2_v is determined uniquely. Then

$$B(b)B(b^{-1})\sigma^2_w = \Lambda(b)A(b)A(b^{-1}) - A(b)A(b^{-1})\sigma^2_v \tag{5}$$

and if we require that the zeroes of B lie on the unit disk, B(b) and σ^2_w are uniquely determined.

Suppose instead that $p \leq q$, and that $(A(b), B(b), \sigma^2_w, \sigma^2_v)$ solve equation (4). Select $\overline{\sigma}^2_v \geq 0$ arbitrarily, and solve the following equation for the order q polynomial $\overline{B}(b)$, constraining its zeros to lie on the unit disk.

$$\overline{B}(b)\overline{B}(b^{-1})\overline{\sigma}^2_w = B(b)B(b^{-1})\sigma^2_w + A(b)A(b^{-1})(\sigma^2_v - \overline{\sigma}^2_v) \tag{6}$$

The result is on ARMA(p,q) model with new parameters $(A,\bar{B},\bar{\sigma}^2_w,\bar{\sigma}^2_v)$ that has the same spectrum as the model with parameters $(A,G,\sigma^2_w,\sigma^2_v)$. By continuity and arbitrariness of $\bar{\sigma}^2_v$ we can find such a new parameter vector in any neighborhood of $(A,B,\sigma^2_w,\sigma^2_v)$, which cannot therefore be locally unique. Q.e.d.

2.4 State Space Representation

The state space representation of the process is most convenient from the point of view of conceptualization, identification, and estimation of the system, and is more flexible mathematically.

We present a representation that is a minor variation of the observable form discussed, for instance, by Brockett [76]. It has been previously used by Mehra [25] and others. For the single input-single output case considered in this thesis, the representation is of minimal dimension. The multi input-multi output generalization presented by Brockett is observable, but not in general minimal. In constructing the representation we reindexed the sequence $\{U_k\}$ in a minor way: $U_{k+r-1} \longrightarrow U_k$. The representation is:

$$\begin{aligned} X_{k+1} &= FX_k + GW_k + CU_k + CR \\ Z_k &= HX_k + V_k \end{aligned} \tag{1}$$

where matrix structures are fixed as follows

$$\begin{aligned} G' &= (G_r, G_{r-1}, \ldots, G_1) \\ H &= (0, 0, \ldots, 0, 1) \\ C' &= (1, 0, 0, \ldots 0, 0) \\ F &= \begin{bmatrix} A_1 & A_2 & \cdot & \cdot & A_{r-1} & A_r \\ 1 & 0 & & & 0 & 0 \\ 0 & 1 & & & 0 & 0 \\ & & \cdot & \cdot & \cdot & \cdot \\ & & & & 1 & 0 \end{bmatrix} \end{aligned} \tag{2}$$

Matrix Structures

We remark that in identification of stationary ARMA models, as discussed, e.g., by Box and Jenkins [3], uniqueness is secured by requiring asymptotic stability and inverse stability, which holds when all roots of B(z) lie within the unit circle. These conditions are also important for identification of our non-stationary state space models, as demonstrated in Appendix 4. The relation between ARMA and state space algebraic forms for these conditions is obtained by equating asymptotic (stationary) spectra for the two representations. The resulting correspondences are shown below.

	ARMA	State Space
Strong Spectral Factor	$\frac{B(z)}{A(z)}$	$\Psi(z) = \sum_{k=0}^{\infty} HF^k Gz^k$
Asymptotic Spectral Density	$\frac{B(z)B(z^{-1})}{A(z)A(z^{-1})} + \sigma_v^2$	$\Psi(z)\Psi(z^{-1}) + \sigma_v^2$
Asymptotic Stability	Roots of A are inside unit circle	Roots of F are inside unit circle
Inverse Stability	Roots of B are inside unit circle	Roots of $\Psi^{-1}(z)$ are inside unit circle

2.5 Formal Extension of SS Model to Include Individual Random Components

In Section 2.4 a state space representation was developed that led to the model specified by Equations (2.4.1), (2.4.2), involving the random term R explicitly. Now the state space representation will be formally expanded to include it. The trend term R may be considered an immutable part of the initial state. Assume the transformation:

$$X_k \longrightarrow \begin{bmatrix} X_k \\ R \end{bmatrix}$$

$$G \longrightarrow \begin{bmatrix} G \\ 0 \end{bmatrix}$$

$$C \longrightarrow \begin{bmatrix} C \\ 0 \end{bmatrix}$$

$$F \longrightarrow \begin{bmatrix} F & 0 \\ 0 & 1 \end{bmatrix} \qquad (1)$$

$$H \longrightarrow (H,0)$$

Then our state space model can be rewritten in the standard form

$$X_{k+1} = FX_k + GW_k + CU_k. \qquad (2)$$

$$Z_k = HX_k + V_k,$$

where the random trend model is formally absorbed into the standard state space form. The advantage is that the Kalman filter estimation technique to be developed can be programmed to accomodate this option with little modification.

2.6 Probability Theory Foundation: Specification of Parameters

Now that a mathematical growth model has been constructed we can discuss the structure of the probability space of trajectories $\{Z_k\}$, laying a foundation for Chapter III, where the ML estimation strategy is presented. (We remind the reader that R^N is the N-dimensional real field while $\mathcal{B}^N$ is the corresponding Borel field.)

We define the sample space $\mathbf{X}$, constructed as the direct product of three subspaces: the <u>population sample space</u> $\mathbf{X}_p$, the <u>environment sample space</u> $\mathbf{X}_w$, and the <u>measurement error sample space</u> $\mathbf{X}_v$.

(1) A point $(X_1, R) \in R^{r+1}$ in the population sample space $\mathbf{X}_p$ represents the initial state of a given subject. The vector is (r+1)-dimensional, representing initial values X_1 for the r-dimensional ARIMA model, and R. We assume that X_1 and R have a joint multivariate normal distribution:

$$\begin{bmatrix} X_1 \\ R \end{bmatrix} \sim N\left(\begin{bmatrix} X_{1|0} \\ 0 \end{bmatrix}, \begin{bmatrix} V_{1|0} & D_R \\ D_R & \sigma_R^2 \end{bmatrix} \right)$$

(2) Points in $\mathbf{X}_w$ represent sequences of stochastic inputs $\{W_k\} \in R^T$, assumed to be i.i.d. normal, with zero mean and variance σ^2_w. (There is an additional deterministic input sequence $\{U_k\}$.)

(3) Points in $\mathbf{X}_v$ represent sequences of measurement errors $\{V_k\} \in R^T$, assumed i.i.d. normal, with zero mean and variance σ^2_v.

The linear model specified by the dynamic equation (2.4.1) and the measurement equation (2.4.2) define $\{Z_k\}$ as a random variable, mapping $\mathbf{X}_p \times \mathbf{X}_w \times \mathbf{X}_v$ to R^T in two stages:

$$\begin{array}{|l|r|} \hline \text{Dynamic Model:} & \mathbf{X}_p \times \mathbf{X}_w \xrightarrow{A,G} \mathbf{X}_x \\ & (X_o, R), \{W_k\} \xrightarrow{A,G} X_k \\ \text{Measurement Model:} & \mathbf{X}_x \times \mathbf{X}_v \xrightarrow{\sigma_v^2} R^T \\ & \{X_k\}, \{V_k\} \xrightarrow{\sigma_v^2} \{Z_k\} \\ \hline \end{array} \qquad (1)$$

In equation (1) the dynamic model maps the initial state and the sequence of inputs $\{W_k\}$ into the sequence of states $\{X_k\}$, which in turn maps into the sequence $\{Z_k\}$. The map is indexed by dynamic model parameters A,G and by the environmental input sequence $\{U_k\}$. By treating sequences as mathematical objects we subsume the actual dynamical relations between X_{k+1} and X_k, k=1,2,...,T-1. Although this formulation ignores interesting parts of the dynamic structure of the process, it focuses attention on the cross-sectional nature of the statistical formulation of our problem.

The probability model is indexed by certain parameters to be identified; these are summarized in Table 1.

Model Component	Description of Parameter	Parameter Vector $\theta = (\theta_I; \theta_{II})$	
		θ_I Stochastic	θ_{II} Deterministic
ARMA	Autoregressive	A	
	Impulse Response	G	
$\mathbf{X}_w$	Environmental Input	σ_w^2	U_k
$\mathbf{X}_v$	Measurement Error	σ_v^2	
$\mathbf{X}_p$	Population Distribution	$V_{1\|0}$	$X_{1\|0}$
		D_R	
		σ_R^2	

Summary of Model Parameters

Table 1

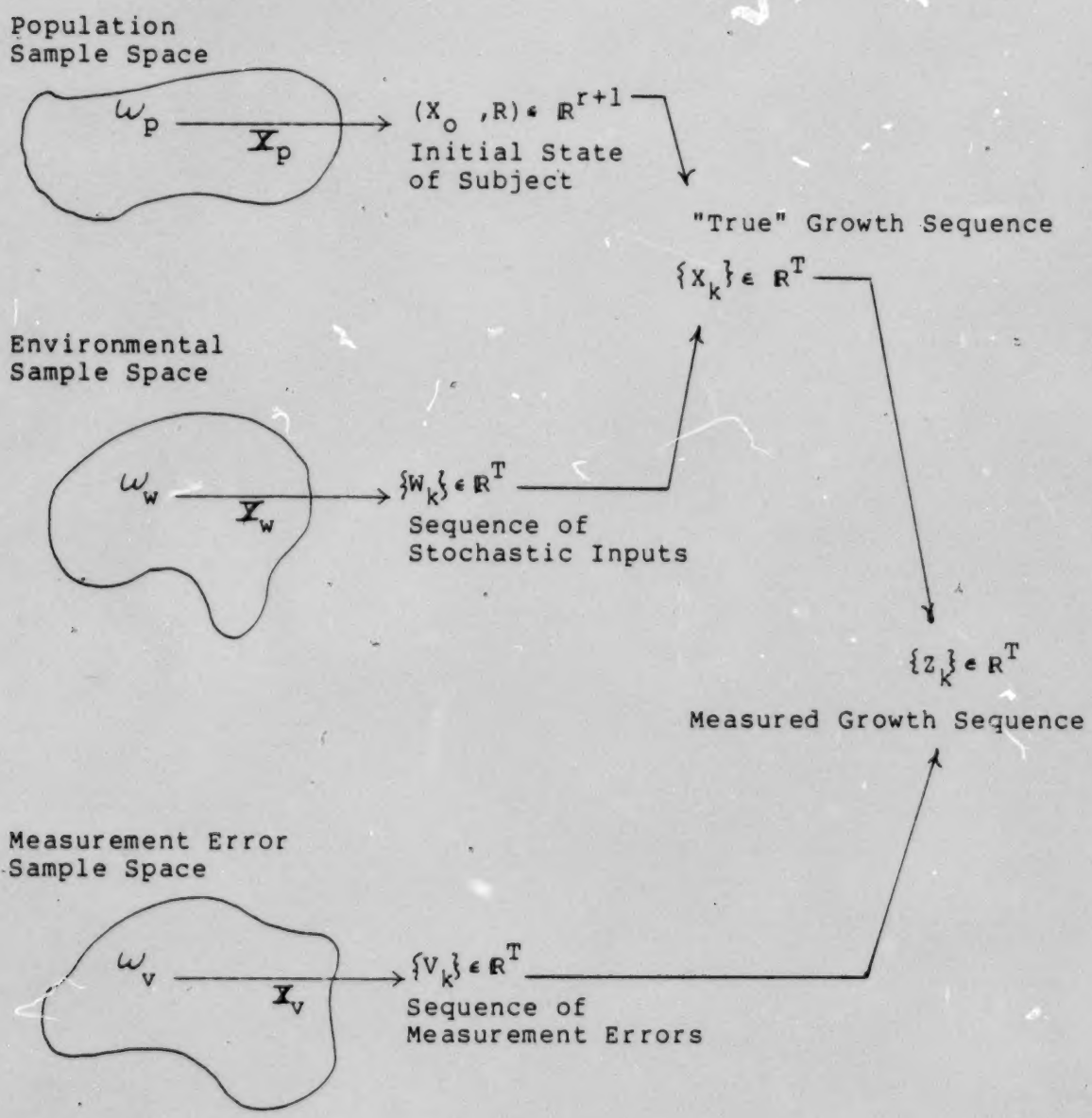


Figure 1

Structure of Random Variable $\{Z_k\} : \mathbb{X}_p \times \mathbb{X}_w \times \mathbb{X}_v \xrightarrow{\theta} \mathbb{R}^T$

Chapter III Model Identification

3.1 Introduction

In this chapter a computational technique is presented for identification of the model presented in Chapter II; a series of theorems regarding identifiability and consistency is presented. The technique is based on numerical maximization of the likelihood function, using the Kalman filter as a computational device. It yields consistent estimates of parameter values, subject to certain identifiability conditions to be discussed. Model structure determination, implementation, and results from practical experience are also discussed. A listing of the computer program is provided in Appendix 8.

3.2 Maximum Likelihood Estimates of Parameters

The problem of model identification is complicated by several features of the model and the data.

(1) The initial state distribution influences the measurement sequence throughout the period of observation, resulting in nonstationary series dependent on initial values.

(2) The problem of random measurement error is always significant in human developmental measurements and must be explicitly considered in the identification strategy.

(3) Series of human developmental measurements often involve irregular measurement points. For instance the Fels data include measurements at half year intervals from 2 1/2

to 6 and full year from 6 to 17, with missing values at 13 and 16. In order to use the data fully such irregularities must be accomodated in the identification strategy.

These facts rule out the use of the estimation techniques usually used, including maximum likelihood methods as discussed, for instance, by Box and Jenkins [3], who assume stationarity and conceptualize the series as originating in the remote past. Instead, we will use exact maximum likelihood estimation techniques, including the distribution of initial conditions parametrically.

The likelihood function is constructed through the so-called "innovations" (Kalman filter) approach. Kalman filter methods for system identification have been discussed by several authors, including Kailath [59], Mehra [26,27,28], Ljung [9,10,11], Caines [12], and Caines and Rissanen [35]. This thesis applies the basic method described by these authors to the cross-sectional case.

We can put the model of Chapter II into the following form, convenient for a unified discussion of the Kalman filter approach.

$$
\begin{aligned}
X_{k+1} &= FX_k + GW_k + CU_k \\
Z_k &= HX_k + V_k \\
X_1 &\sim N(\bar{X}_{1|0}, V_{1|0}) \\
\begin{bmatrix} W_k \\ V_k \end{bmatrix} &\sim \text{i.i.d } N\left(0, \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right) \\
E\, X_1 W_k &= E X_1 V_k = 0
\end{aligned}
\qquad (1)
$$

X_k, G, C, H'	sx1
F	sxs
W_k, U_k, Z_k, V_k	1x1

The method to be presented can be extended to more complicated conditions as follows.

(1) The constant matrices F, G, C, H can be made time varying at little cost in computational complexity.

(2) Nonlinearities can be introduced, with considerable increase in complexity, both computational and theoretical. In particular the measurement process can be made non-linear:

$$Z_k = H(X_k) + V_k$$

This extension would be useful in treating tests with ceiling effects or other curvilinear response characteristics. The computational cost is that the Kalman filter becomes a function of the state estimate and of time k, so that it must be recomputed for each sample, making this extension prohibitively expensive.

(3) The error terms V_k and W_k can be allowed to be correlated, with modest increases in computational complexity.

(4) The scalars W_k, Z_k, V_k can be extended to vector versions. Thus multivariate measurement sequences $\{Z_k\}$ can be considered. Considerable increase in computation complexity results.

Definition of Kalman Filter and Innovations

The linear model described is associated with the following Kalman filter algorithm, originally described by Kalman [83] and discussed in many books including Bryson and Ho [54], Sage and Melsa [53] and Åström [13].

$$
\begin{aligned}
\hat{X}_{k+1|k} &= F\hat{X}_k + CU_k \\
\hat{X}_{k+1} &= \hat{X}_{k+1|k} + K_{k+1}\,(Z_{k+1} - H\hat{X}_{k+1|k}) \\
K_{k+1} &= V_{k+1|k}\,H'(HV_{k+1|k}H' + \sigma_v^2)^{-1} \\
V_{k+1|k} &= FV_kF' + G\sigma_w^2G' \\
V_{k+1} &= (1-K_{k+1}H)\,V_{k+1|k}
\end{aligned}
\tag{2}
$$

Kalman Filter Equations

$$
\begin{aligned}
\hat{X}_{1|0} &= \bar{X}_{1|0} \\
V_{1|0}
\end{aligned}
$$

A Priori Parameters

$\hat{X}_{k+1\|k}$	State Forecast	sxl
$\hat{X}_{k+1}$	Filtered State Estimate	sxl
K_{k+1}	Kalman Filter Gain	sxl
$V_{k+1\|k}$	Covariance of State Forecast	sxs
V_{k+1}	Covariance of Filtered State Estimate	sxs

Names and Dimensions of Newly Introduced Variables

The state forecast and filtered estimate are projections on the space generated by the random variables $\{Z_m; k=1,2,\ldots,k\}$, whose joint distribution is defined by the parameter set $\theta = (F,G,\sigma_w^2,\sigma_v^2,V_{1|0},X_{1|0},\{U_k\})$. Thus

$$\hat{X}_k = E\,(X_k|Z_k,\ldots,Z_1)$$

$$\hat{X}_{k+1|k} = E\,(X_{k+1}|Z_k,\ldots,Z_1).$$

The innovation (at time k+1) $I_{k+1}=Z_{k+1}-H\hat{X}_{k+1|k}$ is a random variable representing the forecast error. The Kalman filter is optimal in the sense that $\hat{X}_{k+1|k}$ minimizes $E(I^2_{k+1}|Z_k\ldots,Z_1)$. Under the additional assumption of normality for all distributions of the model, the sequence $\{I_k;k=1,2,..T\}$ are independently normally distributed:

$$\boxed{\begin{aligned} I_k &= Z_k - H\hat{X}_{k|k-1} \\ I_k &\perp I_m \text{ for } k \neq m \\ P_k &= HV_{k|k-1}H' + \sigma_v^2 \\ I_k &\sim N(0,P_k) \end{aligned}} \qquad (3)$$

Distribution of Innovations

Construction of LF

Let the function $f(Z_1,Z_2,..,Z_T|\theta)$ be the probability of observing $(Z_1,Z_2,\ldots,Z_T)$ given the parameter θ. Then $f(Z_1,Z_2,..,Z_T|\theta)$ can be decomposed using the chain rule for conditional probabilities:

$$f(Z_1,Z_2,..,Z_T|\theta) = \qquad (4)$$

$$f(Z_t|Z_1,Z_2,..,Z_{T-1},\theta)f(Z_{t-1}|Z_1,Z_2,..,Z_{T-2},\theta)\ldots f(Z_2|Z_1,\theta)f(Z_1|\theta)$$

Since $\quad Z_k - E(Z_k|Z_1,\ldots,Z_{k-1},\theta) = Z_k - H\hat{X}_{k|k-1} = I_k,$

and $I_k \perp I_m$ for $m\neq k$, we can compute the likelihood function for a single trajectory $\{Z_k\}$ as a simple product:

$$f(Z_1,Z_2,..,Z_T|\theta) = \prod_{k=1}^{T} \frac{1}{\sqrt{2\pi P_k}} \exp - \frac{I_k^2}{2P_k} \qquad (6)$$

Since trajectories $\{z_k\}$ are i.i.d, across cases the LF for a sample of N trajectories is

$$f(z_1,..,z_T|\theta) = \prod_{i=1}^{N} \prod_{k=1}^{T} \frac{1}{\sqrt{2\pi P_k}} \exp - \frac{I_{ki}^2}{2P_k}, \qquad (7)$$

where I_{ki} is the innovation for case i at time k, and the log likelihood function (LLF) is given (up to an additive constant term) by

$$L(\theta|\{z_k\}) = \sum_{k=1}^{T} \left(\sum_{i=1}^{N} I_{ki}^2/2P_k + \frac{N}{2} \log P_k \right). \qquad (8)$$

Equations (2), (3), and (8) define the log likelihood function recursively as a function of the measurements $\{z_k\}$.

Implementation of Technique

Computational algorithms for minimization require evaluation not only of L, but also its gradient and an approximation to the Hessian matrix. The extension of the equations to provide $\frac{\partial L}{\partial \theta}$ and $\frac{\partial^2 L}{\partial \theta \partial \theta'}$ is derived in Appendix 1. The extension takes the form of an additional set of recursive equations in the gradients of $\hat{X}_{k+1|k}$, K_{k+1}, $V_{k+1|k}$, obtained by differentation, together with formulas for $\frac{\partial L}{\partial \theta}$ and $\frac{\partial^2 L}{\partial \theta \partial \theta'}$. The resulting system of equations, presented in Table 1, was programmed in FORTRAN for the CDC 6400 computer.

Missing Measurements

In the method described, it is assumed that data are available at equal intervals. However, the Fels data analyzed in Chapter IV were recorded at half year intervals from ages 2 1/2 through 6, full year from 6 through 12, and at 14, 15, and 17 years. The Kalman filter method was adapted to this situation by defining the half-year as the time increment, treating missing measurements as present, but with infinite measurement error so that they are effectively removed from the analysis without upsetting the form of the recursion algorithms.

3.3 Consistency of Global ML Estimates

In this section we will show that the ML procedure specified in Table (3.2.1) yields consistent parameter estimates. In two related cases, consistency of ML estimates is well known. They are:

(1) The case of identically independently distributed samples, as discussed by Wald (37). Convergence to true values occurs as the (cross-sectional) sample size approaches infinity.

(2) The case of a stationary, regular time series of correlated samples, as discussed by Caines and Rissanen [35]. Convergence to true values occurs as the (longitudinal) sample size approaches infinity.

The present application does not fit either of these categories. Although we are using the dynamic models used in time series, the number of time points T is small

State Prediction Equations	$\hat{X}_{k+1\|k} = F\hat{X}_{k\|k-1} + FK_{k+1}(Z_{k+1} - H\hat{X}_{k+1\|k}) + CU_k$
Filter Gain Equations	$K_{k+1} = V_{k+1\|k} H'(HV_{k+1\|k}H' + \sigma_v^2)^{-1}$
Variance Equations	$V_{k+1\|k} = FV_kF' + G\sigma_w^2G'$ $V_{k+1} = (1-K_{k+1}H)V_{k+1\|k}$

Kalman Filter Equations

State Prediction Gradient Equation	$\frac{\partial \hat{X}_{k+1\|k}}{\partial\theta} = F(I-K_kH)\frac{\partial \hat{X}_{k\|k-1}}{\partial\theta} + \frac{\partial F}{\partial\theta}(\hat{X}_{k\|k-1} + K_kI_k) + F\frac{\partial K_k}{\partial\theta}I_k$
Filter Gain Gradient Equation	$\frac{\partial K_{k+1}}{\partial\theta} = \frac{\partial V_{k+1\|k}}{\partial\theta}H'P_{k+1}^{-1} - V_{k+1\|k}H'P_{k+1}^{-2}\left\{H\frac{\partial V_{k+1\|k}}{\partial\theta}H' + \frac{\partial\sigma_v^2}{\partial\theta}\right\}$
Variance Gradient Equations	$\frac{\partial V_{k+1\|k}}{\partial\theta} = \frac{\partial F}{\partial\theta}V_kF' + FV_k\frac{\partial F'}{\partial\theta} + \left\{\frac{\partial G}{\partial\theta}G' + G\frac{\partial G'}{\partial\theta}\right\}\sigma_w^2$ $+ F\frac{\partial V_k}{\partial\theta}F' + GG'\frac{\partial\sigma_w^2}{\partial\theta}$ $\frac{\partial V_k}{\partial\theta} = \frac{\partial V_{k\|k-1}}{\partial\theta} - \left\{\frac{\partial V_{k\|k-1}}{\partial\theta}H'HV_{k\|k-1} + V_{k\|k-1}H'H\frac{\partial V'_{k\|k-1}}{\partial\theta}\right\}P_k^{-1}$ $+ V_{k\|k-1}H'HV'_{k\|k-1}P_k^{-2}\left\{H\frac{\partial V_{k\|k-1}}{\partial\theta}H' + \frac{\partial\sigma_v^2}{\partial\theta}\right\}$

Kalman Filter Gradient Equations

Innovation Variance Equation	$P_{k+1} = HV_{k+1\|k}H' + \sigma_v^2$
Likelihood Function	$L = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{T}(I_{ki}^2P_k^{-1} + \log P_k).$

Likelihood Function Equations

Table 1

Log Likelihood Function and Gradient Algorithms

Innovation Variance Gradient Equation	$\frac{\partial P_{k+1}}{\partial \theta} = H \frac{\partial V_{k+1\|k}}{\partial \theta} H' + \frac{\partial \sigma_v^2}{\partial \theta}$
Likelihood Function Gradient	$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{N} \sum_{k=1}^{T} \left\{ - I_k H \frac{\partial \hat{X}_{k\|k-1}}{\partial \theta} P_k^{-1} - \frac{1}{2} I_k^2 P_k^{-2} \frac{\partial P_k}{\partial \theta} + \frac{1}{2} \frac{\partial P_k}{\partial \theta} P_k^{-1} \right\}$

Likelihood Function Gradient Equations

Initial State Estimate	$\hat{X}_{1\|0} = 0$
Initial Variance	$V_{1\|0}$ = Parameters to be determined: See below.
Initial State Gradient	$\frac{\partial \hat{X}_{1\|0}}{\partial \theta} = 0$
Initial Variance Gradient	$\frac{\partial V_{1\|0}}{\partial \theta} = \begin{cases} 1, & \text{For elements corresponding to parameter } V_{1\|0} \\ 0, & \text{Remainder of parameter vector} \end{cases}$

A Priori Distribution Equations

AR Parameters	A	rx1	
IR Parameters	G	rx1	$G_1 \equiv 1$
Process Noise	σ_w^2	1x1	
Measurement Error	σ_v^2	1x1	
Initial State Covariance	$V_{1\|0}$	rxr	with r(r+1)/2 independent elements
Initial State-Trend Covariance	D	rx1	
Trend Covariance	σ_R^2	1x1	

Parameters to be Determined

Table 1 (Continued)

Log Likelihood Function and Gradient Algorithm

and convergence of estimates with large T is not of interest. We are interested rather in convergence as the ensemble size N approaches ∞ .

Definition of Identifiability

It is intuitively clear that a model involving too many parameters is not identifiable when the measurement sequence is short, regardless of the ensemble size. Unless some minimum number of longitudinal time points is available a multiplicity of models explain the data equally well. For instance if the true model were AR(3) we cannot identify all three coefficients using only two data points, and no estimator for these coefficients could be consistent.

Apparently, there is a class of models that are not identifiable given sample sequences of fixed length T, although they are identifiable as $T \longrightarrow \infty$. This fact was considered in constructing the following definition of identifiability.

Definition Let $\{Z_k;\ k=1,2,..,T\}$ be distributed with first and second moments $\mu_T(\dot{\theta})$ and $\Sigma_T(\dot{\theta})$. Let $\dot{\theta}$ be a point in the subset Ω. Then $\dot{\theta}$ is second order identifiable (T,Ω) iff

$$(\forall \theta \in \Omega)\ \mu_T(\theta) = \mu_T(\dot{\theta}),\ \Sigma_T(\theta) = \Sigma_T(\dot{\theta}) \rightarrow \theta = \dot{\theta}\ , \qquad (1)$$

i.e., iff $\dot{\theta}$ is the unique parameter in Ω with image (μ_T, Σ_T). We say that $\dot{\theta}$ is second order locally identifiably (T) iff $\exists$ a neighborhood Ω of $\dot{\theta}$ such that $\dot{\theta}$ is second order identifiable (T, Ω).

We will usually omit the qualification "second order" in discussing identifiability. For the gaussian systems of main interest in this thesis, the qualification is superfluous.

Our definition might be called a structural one, distinguishing it from definitions that rely on the existence of consistent sequences of estimators, such as the definitions of Tse [70,71] and Ljung [10]. Specifically, Tse and Anton [71] offer the following:

Alternative Definition (Tse and Anton) The parameter is said to be identifiable if there exists a sequence of estimates $\{\theta_T; T=1,2,\ldots,\infty\}$ which is consistent in probability.

In this definition it is assumed that only one realization of the process is observed. "Consistency" means convergence of estimators as the length of the observed segment of the process approches ∞. We cannot use this definition directly even though our dynamic model is the same, since our data collection model is that of increasing cross-sectional sample size. Our own definition works for both longitudinal and cross-sectional cases, since it deals directly with the structure generated in the first two moments of the distribution of observations.

The following simple nesting theorem follows immediately from the definition.

Theorem 1 Let the parameter $\dot{\theta}$ be identifiable $(T+1,\Omega)$. Then $\dot{\theta}$ is identifiable (T,Ω).

Proof Let $\theta,\dot{\theta}\in\Omega$ and let $\dot{\theta}$ be identifiable (T,Ω). Suppose that $\Sigma_{T+1}(\theta) = \Sigma_{T+1}(\dot{\theta})$ and $\mu_{T+1}(\theta) = \mu_{T+1}(\dot{\theta})$. Then also $\Sigma_T(\theta) = \Sigma_T(\dot{\theta})$ and $\mu_T(\theta) = \mu_T(\dot{\theta})$. Since $\dot{\theta}$ is identifiable (T,Ω) this implies that $\theta=\dot{\theta}$, i.e., $\dot{\theta}$ is identifiable $(T+1,\Omega)$.

Identifiability and ML(T,$\mathfrak{R}$) Consistency

We will now prove that the our ML(T,$\mathfrak{R}$) estimators are consistent provided θ^0 is identifiable (T,$\mathfrak{R}$). First, ML(T,$\mathfrak{R}$) estimators will be defined precisely.

Definition Let $\theta_n \in \mathfrak{R}$;n=1,2,..., where $\mathfrak{R}$ is a compact region. Then the sequence $\{\theta_n\}$ is ML(T,$\mathfrak{R}$) if $(\forall n)(\forall\theta\in\mathfrak{R})\, L(\mu_T(\theta),\Sigma_T(\theta),\bar{X}_n,S_n) \le L(\mu_T(\theta_n),\Sigma_T(\theta_n),\bar{X}_n,S_n)$, where $\bar{X}_n,S_n$ are sample values for the first and second moments of sequences $\{z_k;k=1,2,..,T\}$, and L is the log likelihood function,

$$L(\mu_1,\Sigma_1,\mu_2,\Sigma_2) = -\frac{n}{2}\{\log|\Sigma_1| + \text{trace}\,\Sigma_1^{-1}\Sigma_2 + (\mu_1-\mu_2)'\Sigma_1^{-1}(\mu_1-\mu_2)\}. \quad (2)$$

The following facts are simple to prove:

$$L(\mu_1,\Sigma_1,\bar{X},S) = L(\mu_2,\Sigma_2,\bar{X},S) \Rightarrow \mu_1=\mu_2,\ \Sigma_1=\Sigma_2 \quad (3)$$

$$L(\mu_1,\Sigma_1,\bar{X},S) \le L(\bar{X},S,\bar{X},S) \quad (4)$$

The proof of consistency depends on the following well known theorem, particularizing the consistency theorem for ML estimators as discussed by Wald [36].

Theorem 2 Consider an infinite sequence of i.i.d. samples of sizes $\{n\}$ from the distribution $N(\mu_T,\Sigma_T)$ of a T-variate random variable. Let $\bar{Z}_n$ and S_n denote sample values for the first and second moments μ_T and Σ_T. Then

(1) $(\forall\bar{Z},S)\ L(\bar{Z},S,\bar{Z}_n,S_n) \le L(\bar{Z}_n,S_n,\bar{Z}_n,S_n)$

i.e., $\bar{Z}_n,S_n$ are ML(T,$\mathfrak{R}^m$).

(2) As $n \longrightarrow \infty$, $S_n \xrightarrow{a.s.} \Sigma_T$

$$z_n \xrightarrow{a.s.} \mu_T$$

i.e., $\{\bar{z}_n\}$, $\{S_n\}$ are strongly consistent

The style of the argument made in the following theorem was originally due to Kendall and Stuart [31]. It has since been extensively used by Åström and Bohlin [67], and by Caines [12]. A similar argument is made in the proof of Theorem 3.4.1.

Theorem 3 Let $\{\theta_n\}$ be ML(T,Ω) and let $\dot{\theta} \in \Omega$ be identifiable (T,Ω). Then $\theta_n \xrightarrow{a.s.} \dot{\theta}$, i.e., the ML$(T,\Omega)$ estimator is strongly consistent.

Proof Let $\dot{\theta} \in \Omega$ and let $\{\theta_n\}$ be ML(T,Ω). Then

$$L(\mu\dot{\theta}), \Sigma(\dot{\theta}), \bar{X}_n, S_n) \leq L(\mu(\theta_n), \Sigma(\theta_n), \bar{X}_n, S_n)$$

$$\leq L(\bar{X}_n, S_n, \bar{X}_n, S_n) \qquad (6)$$

The first inequality follows by definition of ML(T,Ω) sequences and the second from equation (4).

By Theorem 2, $X_n \xrightarrow{a.s.} \mu(\dot{\theta})$, $S_n \xrightarrow{a.s.} \Sigma(\dot{\theta})$ as $n \longrightarrow \infty$. Since L is continuous, the first and third terms of (6) converge almost surely to the common limit $L(\mu(\dot{\theta}), \Sigma(\dot{\theta}), \mu(\dot{\theta}), \Sigma(\dot{\theta}))$.

Therefore the second term of (6) must approach the same limit:

$$L(\mu(\theta_n), \Sigma(\theta_n), \bar{X}_n, S_n) \xrightarrow{a.s.} L(\mu(\mathring{\theta}), \Sigma(\mathring{\theta}), \mu(\dot{\theta}), \Sigma(\dot{\theta})) \quad (7)$$

Since $\theta_n \in \mathcal{R}$ for every n, and $\mathcal{R}$ is compact, $\{\theta_n\}$ has at least one accumulation point. Let $\tilde{\theta}$ be an accumulation point of $\{\theta_n\}$, so that

$$L(\mu(\theta_n), \Sigma(\theta_n), \bar{X}_n, S_n) \xrightarrow{a.s.} L(\mu(\tilde{\theta}), \Sigma(\hat{\theta}), \mu(\dot{\theta}), \Sigma(\dot{\theta})). \quad (8)$$

Comparing (7) and (8) we obtain

$$L(\mu(\tilde{\theta}), \Sigma(\tilde{\theta}), \mu(\dot{\theta}), \Sigma(\dot{\theta})) = L(\mu(\mathring{\theta}), \Sigma(\mathring{\theta}), \mu(\dot{\theta}), \Sigma(\dot{\theta})), \quad (9)$$

so that

$$\mu(\tilde{\theta}) = \mu(\dot{\theta})$$

$$\Sigma(\tilde{\theta}) = \Sigma(\dot{\theta}), \text{ by (3).}$$

Then, by definition of identifiability $(T,\mathcal{R})$, $\tilde{\theta} = \dot{\theta}$. Therefore, there is only one accumulation point $\dot{\theta}$, and $\theta_n \longrightarrow \dot{\theta}$.
Q.e.d.

We have proved that the ML method is consistent provided that certain identifiability conditions are satisfied. Before going on to discuss these conditions we will develop the computational technique further. Then identifiability can be discussed in more specific terms (Section 3.5).

3.4 Consistency of Stepwise ML Estimates

In this section we discuss consistency of estimates obtained from a stepwise procedure to be described.

We are interested both in mean developmental curves for a group of individuals and in deviations of individual curves from the mean. The deviations contain stochastic information about the development process, while the mean curve contains information about potential treatment effects and other systematic influences. In this section we separate the problem into steps representing these two statistical aspects, in order to simplify numerical analysis.

We begin by subtracting cross-sectional sample means for each time point from individual values to obtain deviations from the sample mean, removing trends in the data. The resulting deviations satisfy the equations

$$\boxed{\begin{aligned} (X_{k+1} - \bar{X}_{k+1}) &= F(X_k - \bar{X}_k) + G(W_k - \bar{W}_k) \\ Z_k - \bar{Z}_k &= H(X_k - \bar{X}_k) + V_k - \bar{V}_k \end{aligned}} \qquad (1)'$$

By an obvious change of notation, equations (1)' are rewritten as

$$\boxed{\begin{aligned} X_{k+1} &= FX_k + GW_k \\ Z_k &= HX_k + V_k \end{aligned}}, \qquad (1)$$

where $EX_k = EZ_k = 0$. The corresponding equations for the cross-sectional mean are:

$$\boxed{\begin{aligned} \bar{X}_{k+1} &= F\bar{X}_k + G\bar{W}_k + CU_k \\ \bar{Z}_k &= H\bar{X}_k + \bar{V}_k \end{aligned}} \qquad (2)$$

The filter equations can be partitioned in the same manner.

$$\begin{aligned} \hat{X}_{k+1|k} &= F\hat{X}_k \\ \hat{X}_{k+1} &= \hat{X}_{k+1|k} + K_{k+1}(Z_{k+1} - H\hat{X}_{k+1|k}) \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{\bar{X}}_{k+1|k} &= F\hat{\bar{X}}_k + CU_k \\ \hat{\bar{X}}_{k+1} &= \hat{\bar{X}}_{k+1|k} + K_{k+1}(\bar{Z}_{k+1} - H\hat{\bar{X}}_k) \end{aligned} \tag{4}$$

However, the two parts of the problem remain linked through the likelihood function:

$$\begin{aligned} L &= L_I + L_{II} \\ L_I &= \sum_{i,k} (Z_{k+1} - HX_{k+1|k})^2 P^{-1}_{k-1} + \sum_k \ln P_{k+1} \\ L_{II} &= N \sum_k (\bar{Z}_{k+1} - H\bar{X}_{k+1|k})^2 P^{-1}_{k+1}. \end{aligned} \tag{5}$$

If we define θ_I and θ_{II} by

$$\begin{aligned} \theta_I &= (A, G, \sigma_w^2, \sigma_v^2, V_{1|0}) \\ \theta_{II} &= (X_{1|0}, \{U_k\}) \\ \theta &= (\theta_I, \theta_{II}) \end{aligned}, \tag{6}$$

then $L_I = L_I(\theta_I)$ while $L_{II} = L_{II}(\theta_I, \theta_{II})$, and a global maximum of $L(\theta_I, \theta_{II})$ can be found by computing the function

$$L_{II}(\theta_I, \hat{\theta}_{II}(\theta_I)) = \max_{II} L_{II}(\theta_I, \theta_{II}), \tag{7}$$

then finding

$$L(\theta_I, \theta_{II}) = \max_{\theta_I} [L_I(\theta_I) + L_{II}(\theta_I, \hat{\theta}_{II}(\theta_I))]. \tag{8}$$

Although this technique is feasible (Appendix 6) there are high computational costs in computing the function defined by (7), and a simpler procedure was implemented. First, $L_I(\theta_I)$ is maximized over θ_I,

$$L_I(\hat{\theta}_I) = \max_{\theta_I \in \Omega_I} L_I(\theta_I), \tag{9}$$

and then the resulting estimate is used in estimation of θ_{II}:

$$L_{II}(\hat{\theta}_I, \hat{\theta}_{II}) = \max_{\theta_{II} \in \Omega_{II}} L_{II}(\hat{\theta}_I, \theta_{II}). \tag{10}$$

Such a stepwise procedure does not yield the global maximum for (θ_I, θ_{II}) but does yield consistent estimators, as shown in the proof of the following theorem.

Theorem 1 Let θ be identifiable (T, Ω). Then the sequential estimation strategy specified in (9), (10) is strongly consistent.

Proof Denote by $\hat{\theta}_I, \hat{\theta}_{II}$, the stepwise estimates defined in (9), (10) and by $\hat{\hat{\theta}}_I, \hat{\hat{\theta}}_{II}$ the globally maximizing estimates, leaving sample size implicit.

First we will show that $\hat{\theta}_I$ is strongly consistent. We need only observe that, asymptotically, L_I is the likelihood function for the reduced problem $\theta_I = (A, G, \sigma_w^2, \sigma_v^2, V_{1|0})$, $\theta_{II} = 0$. Clearly, by definition (3.3.1), θ_I is identifiable (T, Ω) in the reduced model. Therefore Theorem 3.3 applies and $\hat{\theta}_I$ is strongly consistent.

Now we observe that

$$L(\hat{\theta}_I, \hat{\hat{\theta}}_{II}) \leq L(\hat{\theta}_I, \hat{\theta}_{II}) \leq L(\hat{\hat{\theta}}_I, \hat{\hat{\theta}}_{II}), \tag{11}$$

where the first inequality follows from Equation (10) and the second by definition of the global maximum. Since $\hat{\theta}_I$ and $\hat{\hat{\theta}}_I$ are both consistent estimators of $\dot{\theta}_I$, and since L is continuous, the first and third terms of (11) approach the common limit $L(\dot{\theta}_I, \dot{\theta}_{II})$, and therefore so does the second term. Therefore $(\hat{\theta}_I, \hat{\theta}_{II})$ are asymptotically global ML estimators. By identifiability (T,Ω) and another application of Theorem 3.3.3, they are consistent. Q.e.d.

In one particular case, estimates obtained by the sequential procedure (9), (10) are in fact globally ML for finite sample size n.

Theorem 2 Let the admissible region Ω_{II} for θ_{II} be R^T so that the elements of $\theta_{II} = (X_{1|0}, \{U_k\})$ are unconstrained. Then the estimates $(\hat{\theta}_I, \hat{\theta}_{II})$ equal the ML estimates $(\hat{\hat{\theta}}_I, \hat{\hat{\theta}}_{II})$.

Proof It is shown in Appendix 6 that L_{II} (defined in Equation (5)) can be minimized to zero for any value of θ_I, so that the maximum of $L(\theta_I, \theta_{II}) = L_I(\theta_I)$ is global. Q.e.d.

In most cases it is desired to constrain $\{U_k\}$, since hypotheses about systematic environmental effects take the form of such constraints. In Appendix 6 we consider the class of constraints $(X'_{1|0}, U')A'=0$, where $U'=(U_1, U_2, \ldots, U_{T-r})$, and A is a rectangular SxT matrix of rank equal to S, the number of constraints. In the main text, however, we will focus on the zero mean problem, i.e., on identifying θ_I by performing the maximization indicated in (9).

3.5 Model Identifiability $(T,\mathcal{L})$

We have shown that our estimation procedure is consistent, conditional on model identifiability. In this section we discuss the conditions for identifiability and present a series of relevant theorems.

The following numbers of independent parameters are involved in the vector θ_I, for the model of order r.

	Number of Parameters
A	r
G	r-1
$V_{1\|0}$	r(r+1)/2
D_R, σ^2_R	r+1
σ^2_w	1
σ^2_v	1
Total	$2r+1 + \frac{(r+1)(r+2)}{2}$

Table 1. Numbers of Independent Parameters

Since the number of independent elements in the sample covariance matrix S equals T(T+1)/2, and the parameter vector θ_I must be recovered as a function of Σ_T, the following simple theorem is obtained by counting parameters.

Theorem 1 A necessary condition that the parameter vector θ_I be identifiable $(T,\mathcal{L})$ is that T equal or exceed T* from the following table.

Parameter Vector	Model Order	Sequence Length Time Points Required (T*)
$(A,G,\sigma_w^2,\sigma_v^2,V_{1\|0})$	r	r+2
$(A,G,\sigma_w^2,\sigma_v^2,V_{1\|0},D_R,\sigma_R^2)$	1	3
	r>1	r+3

Table 2. Sequence Lengths Required for Identifiability

In general, the sequence lengths specified in Table 2 are not sufficient for identifiability. An essential requirement for sufficiency is that the model would be disequilibrium, in a sense made precise by the following definition.

Definition The model is said to be in full rank disequilibrium when $|\Sigma_r - \tilde{\Sigma}_r| \neq 0$, where Σ_r is the covariance matrix of $\{z_k; k=1,2,..,r\}$ and $\tilde{\Sigma}_r$ is the covariance matrix of $\{z_k; k=2,3,..,r+1\}$.

However, this definition is often not satisfied; in particular, stationary systems are obviously not in full rank disequilibrium. Heuristically, the covariance matrix of observed stationary sequences is of Toeplitz form, and contains less information about the model than does a covariance matrix of more general form. Thus when the model is not in complete disequilibrium, longer sequences are required. We will need the following definition in formulating conditions for identifiability.

Definition (Kalman) Let F be an r x r and G an r x s matrix. Then (F,G) is controllable iff $(F^{q-1}G, \ldots, F^2G, FG, G)$ has rank r.

The following theorems, which make conditions for identifiability precise, are proved in appendices 4 and 5.

Theorem 2 Consider the reduced model with $\theta=(A,G,\sigma^2_w, \sigma^2_v, V_{1|0})$ and $D_R=\sigma^2_R=0$. Let $\mathcal{R}$ be the set defined by

(1) F is nonsingular
(2) F is asymptotically stable
(3) $G_1>0$
(4) $(\sum_{k=0}^{\infty} HF^kGz^k)^{-1}$ is analytic for $|z|<1$, i.e., θ is inverse stable

Then if $\mathring{\theta}\in\mathcal{R}$ is in full rank disequilibrium, $\mathring{\theta}$ is identifiable $(r+2,\mathcal{R})$. If $(F,V_{1|0}H')$ is controllable, or if the model is is stationary, then $\mathring{\theta}$ is identifiable $(2r+1,\mathcal{R})$.

Theorem 3 Let $\theta_I = (A,G,\sigma^2_w,\sigma^2_v,V_{1|0},D_R,\sigma^2_R)$ and let $\mathcal{R}$ satisfy the conditions of Theorem 2. The $\mathring{\theta}_I\in\mathcal{R}$ is identifiable $(2r+2,\mathcal{R})$.

Note that when the model is stationary, the condition that $(F,V_{1|0}H')$ is controllable is equivalent to the condition that (F,G) is controllable, assured in our minimal representation when F is nonsingular. The requirement of nonsingular F is essential in Theorem 2. These points are discussed in Remarks 1 and 2, following Lemma A4.3 in the appendix.

3.6 Local Identifiability and the Hessian Matrix

The matrix of second derivatives (Hessian) of the log likelihood function plays an important role in numerical optimization. For instance, the Hessian (or an approximate estimate) is used in second order methods such as the Newton-Raphson or Fletcher-Powell algorithms and the efficiency of convergence is

crucially dependent on its properties. In this section we cite connections between local identifiability and the rank of the Hessian matrix. Proof for results to be stated are presented in Appendix 7.

Let us denote by $[L_{ij}]$ the matrix $\left[\frac{\partial L}{\partial \theta_i \partial \theta_j}\right]$, and by $\mathcal{R}$ the set of points that are regular in a sense defined in Appendix 7. Let $\mathcal{R}^C$ denote the complement of $\mathcal{R}$. Then the following theorems are true.

Theorem 1 Let $\mathring{\theta} \in \mathcal{R}$. Then $\mathring{\theta}$ is locally identifiable (T) if and only if, as $n \to \infty$,

$$[L_{ij}] \xrightarrow{a.s.} M>0. \tag{1}$$

Theorem 2 The point set $\mathcal{R}^C$ is of Lebesgue measure zero, so that Theorem 1 is "generically" true.

Theorem 1 extends the work of Tse [70], who proved that the rank condition on $[L_{ij}]$ is sufficient for local identifiability (without the regularity qualification). In constructing our proofs we have been guided by the work of Fisher [60], who discovered the importance of the regularity qualification in an econometric linear model. Our contribution is to extend his methods to our own dynamic model and to relate his results involving the Jacobian of transformation to the Hessian.

Effects of Overfitting

In the parameter estimation procedure, the Hessian matrix is estimated as part of the optimization algorithm. Theorem 1 implies that the estimate of the Hessian matrix must tend towards singularity when $\mathring{\theta}$ is not locally identifiable and conversely that when the Hessian is near-singular, it is probable that $\mathring{\theta}$ is not locally identifiable. Second

order optimization methods that rely on inverting an estimate of the Hessian matrix are likely to be unstable when the parameter is not identifiable. This has been observed repeatedly in practice.

The most common cause of unidentifiability is overfitting the model order, i.e., one assumes that the order is r but in fact $A_r = 0$. Except for the set of irregular points, the following is true.

$$A_r = 0 \iff |F| = 0 \iff \theta \text{ is not identifiable} \iff \left| \left[\frac{\partial^2 L}{\partial\theta_i \partial\theta_j} \right] \right| \xrightarrow{a.s.} 0$$

The first implication follows from consideration of the characteristic polynomial. The second was discussed in Remark 1 following Lemma A4.3, and the third follows from Theorem 1. In estimating the structure of a given model the bulk of computer time and money is spent in futile pursuit of essentially unidentifiable models. The results of this section give us theoretical justification for abandoning such pursuits.

3.7 Structure Determination

There are few essential differences between the problems of structure determination encountered in cross-sectional problems and more conventional longitudinal problems. Structure determination is carried out with a mixture of hypothesis testing techniques and *a priori* knowledge of the systems considered. It is a partly subjective process. We do not believe that a single criterion for model identification can be universally valid. (The criteria founded in information theory proposed by Akaike [32] and Rissanen [45] are theoretically very interesting but do not necessarily optimize the right functions for particular

problems, especially where a priori knowledge is available, or where actions are to be based on the results. Therefore we employ more traditional forms of hypothesis testing. This section will discuss a few specifics, grounded on experience in applying our techniques to identify the parameter vector θ_I.

The principal structure parameters are model order r and the presence/absence of trend effects. Model order and trend/no trend establish a set of structures of increasing complexity. The maximum complexity allowable is defined by identifiability theorems stated earlier in this chapter. For very short measurement sequences, e.g., NDCS data, this places practical limits on the models to be considered.

In practice one estimates models of increasing complexity, using a computer program to be described, until clear cut overfitting occurs. Then certain statistical tests are performed for candidate models. For each structure tried, the information available includes the following (obtained separately for each structure):

(1) Fisher Information Matrix (Hessian) obtained at the maximizing parameter value.

(2) The record of iterations, including step-by-step LLF value, gradient, and Hessian matrix.

(3) Maximum LLF obtained.

These data are used in the following ways to select a model structure.

(1) Often, models do not converge even after long and expensive runs involving hundreds of iterations. Such models are nearly always associated with Hessian matrix estimates that are near singular. On the basis discussed in Section 3.6 these model structures are discarded as unidentifiable.

(2) The Hessian matrices for convergent models are examined. The Hessian is inverted and Cramer-Rao estimates of parameter variances and covariances are inspected. Near unidentifiability is manifested by very large parameter variance estimates and near perfect correlations among different estimates, indicating overfitting. The Fels data analysis includes examples of the use of this criterion, used to justify rejection of certain models from further consideration.

(3) Models that survive rejection by either of these criteria are analyzed using the likelihood ratio test. Each model estimation results in a maximum value for the LLF under the structure assumed. Asymptotically the null hypothesis that the simpler of two models is correct is tested by comparing twice the LLF to points on the chi-square distribution in numbers of parameters. This test is closely related to the Akaike AIC criterion [32].

The following upper bound can be set for the LLF, based on (3.3.4).

$$\text{LLF} \geq \frac{N}{2} (T + \log \det S) \qquad (1)$$

This bound may sometimes be used to avoid numerical estimations of overfitted models. Since these runs are the most resource-consuming (they are usually associated with near-singular Hessian matrices) the advantage can be substantial.

3.8 Implementation of Methods

A computer program was written to implement our identification technique and was debugged using simulated data. Subsequently hundreds of runs were made with real data. It became apparent that the major problem with implementation lay in the numerical maximization method.

The first methods tried included the Fletcher-Powell and modified Newton-Raphson methods. Both of these second order methods tended toward instability. In fact convergence was never obtained. The source of the problem was determined to be the complicated structure of the LF surface, partly caused by near-unidentifiability. Only very near the minimum was the surface sufficiently well represented by the second order approximation to obtain convergence. A further difficulty was that only an approximate value could be computed for the Hessian.

A great variety of methods was tried experimentally, first with simulated data and then with the Fels data. Iterations turned out to be expensive so that efficiency was important. The method finally settled upon was as follows:

(1) Gradient method with step size multiplier and step size maximum set as parameters, followed by:

(2) Newton-Raphson method, with actual step size set to a fraction of N-R calculated step. Fractions were variable in steps from .03 to 1.0.

Most minimizations took several runs and parameter readjustments. Much was learned about these practical problems in analyzing a substantively important real data set, the Fels Longitudinal IQ data, to be described in the next chapter.

Although the numerical methods finally used for the Fels data were satisfactory in the sense that convergence was finally obtained, there is room for improvement. In a closely related problem not reported in this monograph, namely an application of the Kalman method to ML estimation of variance components, the Fletcher-Powell program VA06A was successfully employed. Fletcher [58] claims that VA06A is guaranteed to converge, and in fact it has not failed in the approximately 50 trials made. In the case of

the variance components program it was found useful to reparameterize variances by their square roots, which may be allowed to vary unconstrained by non-negativity conditions. This change appears worth experimental implementation.

Note that there is no assurance that the local minimum found by the algorithm is global. Except when parameter sets have very low dimensionality there is no practical way to obtain an overview of the general shape of the LLF surface. Multiple local minima were found for each of the two data sets analyzed in this thesis.

3.9 Prediction Error Formulation

Stationary nonlinear prediction error (PE) estimators have been studied by Ljung [9,10], Caines [12], and Caines and Ljung [87]. In the stationary case the ML estimator is a special case of PE estimate, i.e., one which minimizes a scalar function (with specific properties stated by Caines and Ljung) of the matrix

$$Q = \sum_{k=1}^{T} I_k R_k I_k',$$

where I_k are (vector) innovations and R_k are positive definite matrices.

The PE estimation method is easily extended to the cross-sectional case by summing over replications as well as time steps, so that

$$Q = \sum_{i=1}^{N} \sum_{k=1}^{T} I_k R_k I_k'.$$

In our application I_k are scalars and one minimizes Q directly.

In the stationary case, the ML technique minimizes the sum

$$\sum_{i=1}^{N} \sum_{k=1}^{T} I_k^2,$$

so that the ML method is a special case of PE estimator, with $R_k=1$. The nonstationary case is more complex.

While ML estimators minimize the sum

$$\sum_{i=1}^{N} \sum_{k=1}^{T} I_k^2 P_k^{-1}$$

for fixed P_k, and are thus PE in that sense, both I_k and P_k are functions of the parameter θ, and we actually minimize

$$\sum_{i=1}^{N} \sum_{k=1}^{T} [I_k^2 P_k^{-1} + \log P_k].$$

Thus the ML estimate is <u>not</u> a PE estimate although the two are closely related.

We have adhered to ML estimates throughout this chapter for the sake of simplicity. However, it seems a promising avenue of research to investigate general PE estimators in the cross-sectional nonstationary case, and also to investigate quality of fit when the true system can only be <u>approximated</u> by the family of models considered. The stationary, longitudinal, case has been studied by Caines [88].

Chapter IV Analysis of Fels Institute Longitudinal Data

4.1 Introduction

In this chapter we apply methods developed in earlier chapters to the analysis of IQ development in 80 subjects over a period of 15 years. These data were made available to us by the Fels Institute and the particular courtesy of Drs. Alex Roche and Robert B. McCall. As cross-sectional time series of medium length (17 points), these data are ideally suited to our methodology, and are of exceptional interest substantitively. We appreciate the assistance and cooperation received from Drs. Roche and McCall.

This work is closely related to that of McCall, Appelbaum and Hogarty as reported in the 1973 SRCD monograph Developmental Changes in Mental Performance [7]. The discussion most relevant to our analysis is given in Chapter V, "Patterns of IQ Change by Age". The 80 subjects discussed there are the same ones we will discuss, and our analysis may be regarded as an extension of the work of McCall et al. For this reason we would recommend to the substantively oriented reader that he begin by reading this important SRCD monograph. We will discuss the characteristics of the data only in summary form.

A total of 80 subjects, 38 males and 42 females, 2 black and 78 white, born between April 1930 and July 1938, is included. Subjects with complete data were given Stanford-Binet IQ assessments at ages 2-1/2, 3, 3-1/2, 4, 4-1/2, 5, 5-1/2, 6, 7, 8, 9, 10, 11, 12, 14, 15 and 17 years, customarily within a week of the child's birthday or half-birthday. Two versions (1916 and 1937) of the Binet were used, employing the Pinneau correction on Forms L and M of the 1937 revision. In order to provide a common scale, 1916 test were rescaled

to match means and standard deviations for Fels subjects of the same age on the 1930 test. Nearly every subject had some missing data. Average of 14 (possible 17) measurements are available for each subject. Missing data were imputed by linear interpolation. In 17 cases there was one instance in which two consecutive cases were missing and in 1 case three consecutive cases were missing. No more severe runs of missing data occcured.

These data were analyzed as though they were repeated measurements of the same phenotypic human trait through the entire age range of the data. Since the test protocols themselves differ from age to age, beginning with mixtures of cognitive and psychomotor items (age 2-1/2) and progressing to almost total concentration on cognitive skills and verbal ability (age 17), one may question this assumption. Furthermore, scores are scaled in accordance with a specific developmental model of cognitive growth and do not pretend to assess absolute intelligence. These problems are significant but they are beyond the scope of this thesis. Our approach to the data makes no assumptions that are not already implicitly or explicitly in the research of McCall, Appelbaum, and Hogarty. Our results are neither more nor less generalizable than the work of these authors.

This work differs from that of McCall et al. in that auxiliary measures of environment or genotype play no role in the analysis. These measures (except for sex) were not made available to us by Fels and the analysis does not require them. Environmental and genotypic influences enter the analysis as statistical distributions whose parameters are to be estimated. Each subject (along with his entire developmental trajectory) is considered as a sample from such an underlying sample space. This approach does not allow separate consideration of genotypic and environmental influences. Thus there is no substantial relevance in our results to the current "nature-nurture" controversy.

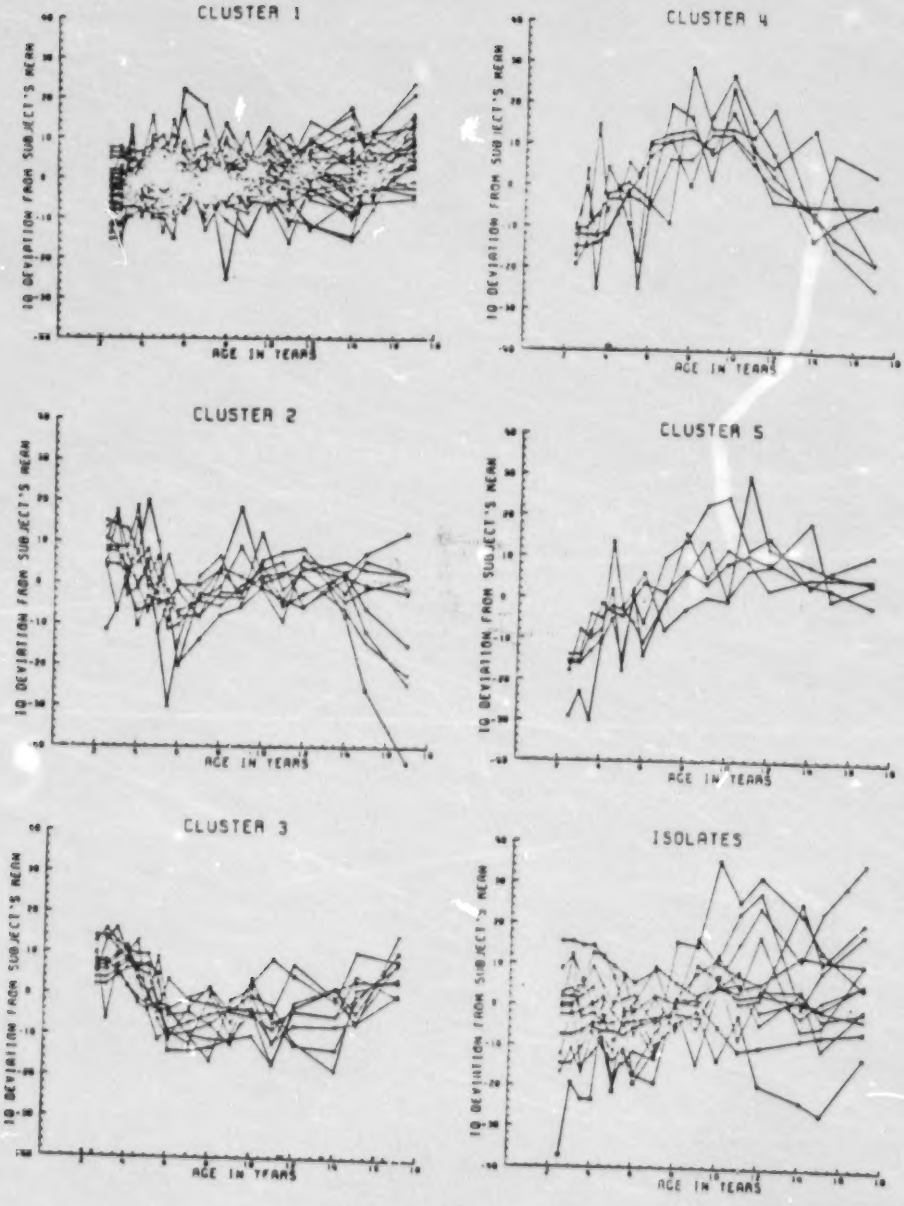


Figure 1

Clustered Trajectories of Fels Longitudinal Sample

Reproduced from McCall et al. [7], p.46, by permission of the Society for Research in Child Development

4.2 Conclusions of McCall et al. and Previous Researchers

At the risk of rather drastic oversimplification we will summarize some of the previous results cited in McCall et al., and most relevant to the present study:

(1) The Fels analyses shows that the concept of IQ scores as fallible measures of stable traits appears false for these subjects, whose IQ's shifted an average of 28.5 points between 2-1/2 and 17 years of age. These shifts are apparently not a result of shifting content of IQ tests with age. (This was demonstrated by Sontog et al. [47] who showed that gainers and losers did not differ markedly in their ability to pass one set of items versus another.)

McCall et al. specifically deny that high correlational stability merits the interpretation that IQ scores are largely unchanging throughout life. Individual trajectories from the Fels data support McCall's contention.

(2) McCall et al. find systematic patterns of growth, i.e., they find it possible to classify most growth trajectories into clusters that are relatively alike in general growth profile within cluster but dissimilar across clusters (Figure 1). Although clusters were defined objectively by analysis of fitted parameters, it is possible to assign most trajectories subjectively as follows:

Cluster 1	36 subjects	Slightly rising pattern through entire series.
Cluster 2	9 subjects	Early decline (4-6 years) and slight recovery, followed by heterogeneous patterns.
Cluster 3	10 subjects	Decline followed by recovery.

Cluster 4	7 subjects	Rise followed by decline.
Cluster 5	5 subjects	Rise followed by shorter decline than in Cluster 4.
Isolates	13 subjects	Unassigned to cluster.

McCall et al. found parental correlates with these patterns, reflecting attitude towards acceleration of child development and towards severity of punishment. These results suggested that there is a stable component of environment, varying from family to family, that "drives" changes in IQ of developing individuals.

(3) McCall et al. are careful to distinguish between general level and longitudinal profiles of IQ. They believe that these two variables may be independent and treat them separately, correcting trajectories for mean IQ for the purposes of profile analysis. Data on the correlates of IQ change are much less clear than on correlates of IQ general level. In the Fels data there is not much evidence of correlation of change with presence or sex of younger siblings or family size (Sontog et al., [44]). Sontog found relations with personality correlates such as independence, competitiveness, self initiation and agressiveness. Most importantly, he found that parental accelerational attempt, and an emphasis on rational and democratic discipline, are correlated with IQ change, although data are sparse.

Our interpretation of the previous Fels research results supports the notion that IQ develops as the sum of a series of correlated increments each of which is influenced by environmental input. The Fels data suggest that IQ level and IQ change are substantively different constructs. It appears that initial IQ level may be much more highly correlated with parental characteristics (an imperfect genotypic measure) than is IQ change. On the other hand IQ increments appear to be more influenced by environment provided than by other characteristics. This is in general agreement with the so called overlap hypothesis discussed, for instance, by Bloom [8].

4.3 Hypotheses of this Study

Nearly all researchers agree (see [8], p.71) that the environment influences the development of "intelligence" as measured by IQ. Unfortunately, measures of environment are exceedingly crude. For instance socio-economic status, occupation and education are often used in research as indexes of environment. These may although they are often amended by attitudinal surveys of parental child rearing behavior. Such studies ordinarily show correlations with IQ levels and IQ increments, but such measures are obviously poor as indexes of the actual quality of environment for cognitive development. Research in cognitive development has been held back substantially by the apparent inability to measure precisely the most important causal variables, although there is now considerable research conducted in interaction processes, and the situation is likely to improve.

When environment is studied as an interaction process, it appears more complex and variable than indicated by the traditional environmental variables. The important features of the environment, especially regarding relationships with other individuals inside and outside the home, are in continual flux. We should therefore expect to observe response to these fluctuations, manifested as longitudinal fluctuations in cognitive growth.

Most previous studies assumed that stable features of the environment prevail, and that these stable features are (imperfectly) measured by family status variables. We also assume that the environmental input has such a stable component, exerting a steady influence throughout the development process, but in addition, that there is a stochastic part representing fluctuations in effective quality of the environment. Such a stochastic influence would cause additional variance in the development process, unaccounted for in traditional approaches, and is not to be

confused with measurement error. The existence of this source of variance is the essence of the first hypothesis listed below. Of course if the hypothesis is not true then our approach to the data is unnecessarily complicated, and Bryk's assumption that a "deterministic position is conceptually adequate" would be supported.

The remaining hypotheses are more purely technical in nature. H2 and H4 provide alternative potential explanations for the large changes in IQ that are observed to occur.

H1 The Fels data are parsimoniously explained by a state space model of low order from the family discussed in Chapter II, with white noise stochastic input $\{W_k\}$.

$$\boxed{\begin{aligned} &H1: \sigma_w^2 > 0 \\ &\quad EW_jW_k = \delta_{jk}\sigma_w^2 \end{aligned}}$$

H2 The data are described by an unstable model, predicting substantial and persisting changes in IQ, in accordance with the conclusions of McCall et al.

$$\boxed{H2: \quad (\exists\lambda)\,|\lambda| = 1,\ |F-\lambda| = 0}$$

H3 Data from ages 2-1/2 - 12 support the same model as those from 12-17.

$$\boxed{H3: \quad A,G \text{ are constants.}}$$

H4 There are subject-to-subject trend effects, supporting McCall's cluster classification. Trend and initial IQ are positively correlated.

$$\boxed{\begin{aligned} &H4: \quad \sigma_R^2 > 0 \\ &\qquad HD_R > 0 \end{aligned}}$$

4.4 Identification of System

The Fels data were identified in a sequence of computer runs involving different model structures. Each run resulted in identification of maximum likelihood estimates for parameters in the structure considered. In addition to these estimates, an estimate of the Fisher Information matrix and of the minimum value of the likelihood function were obtained. These quantities were used in hypothesis tests to determine model structure, and to test hypotheses H1-H4.

Three model structures were considered. They are:

I. First order model without trend-input.

$r = 1$
$\sigma_R^2 = 0$ (H4 false)

II. Second Order model without trend-input.

$r = 2$
$\sigma_R^2 = 0$ (H4 false)

III. First Order model with trend-input.

$r = 1$
$\sigma_R^2 > 0,\ D_R > 0$ (H4 true)

Three Data Sets involving different subsets of measurements were considered. They are:

I. The 10 annual measurements from 3 to 12 years are used and other data ignored.

II. The 10 annual measurements from 3 to 12 (Data Set I) are used along with measurements at 14, 15 and 17 years, for a total of 13 measurements.

III. All 14 measurements from 2-1/2 to 12 years are used and data from 14, 15, and 17 years ignored. The model time step is the half year.

Data set III strengthens data set I to achieve more statistical power over the same age range. Only a few runs, unrecorded here, were made with all 17 data points, for economy, and because of the disproportionate weighting given earlier ages in the full data set (eight measures in the first four year, three in last four years). Computer runs were roughly proportional in length to the number of time steps so that data set III requires three times the time of data set I, which was used for most of the work.

A total of seven computer runs play important roles in the identification process. Each "run" actually consisted of a number of separate jobs involving parameter readjustment to obtain convergence of the minimization. In several cases runs were begun using 25 rather than 80 replicates in order to locate $\hat{\theta}$ approximately but all results quoted here pertain to the full sample of 80 replicates, with missing data inputed. Identification results are summarized in Tables 1 and 2 and in Figure 1.

Discussion of Individual Runs

Data Set I

The first three runs used Data Set I for economy in making a preliminary identification, and led to rejection of Model II, for the following reasons.

(1) The Hessian was near singular and the model nearly unidentifiable, indicating overfitting (see Section 3.6). Convergence was secured only by using the gradient method.

(2) The ML estimate obtained was not significantly different from the ML estimate for model I.

Models I and III were both acceptable, although the likelihood ratio test results in acceptance of Model I for significance levels $p<.20$. The parameter for Model I is unstable and that for Model III stable, so that in terms of the hypotheses we have the choice between (Model I) H2 true, H4 false and (Model III) H2 false, H4 true. Data sets II and III provide more powerful tests of the alternatives.

Data Set II

Data set II differs from data set I by the presence of three additional points at 14, 15, and 17 years of age. Using these data, Model I converged to a marginally stable model ($A=.96\pm.01$). The hypothesis that values for A using data sets I and II are equal was rejected (two sided t-test, $t=2.6$, $DF \rightarrow \infty$).

Model III converged (nearly) to the same parameter found for Model I, with $\sigma^2_R=.01$, not significantly different from 0. A near stationary point was located near the point found earlier (Data Set I), which led to considerable difficulty in securing convergence. Since Data Sets I and II lead to different results for either Model I or III, H3 is rejected, and validity of models will be limited to ages 12 and under.

Data Set III

The final two runs (Models I, III) were made with Data Set III, which includes every available point for age 12 and under. Both runs led to parameter estimates that

4.0/A/11--s.s.

Run	Data Set	Time Points	Number of Measure-ments	Model	Minumum LLF	Number of Parameters	Hessian	Stability
1	I	10(Annual)	10	I	2088.3	4	Nonsingular	Unstable
2	I	10	10	II	2084.15	7	Near singular	Unstable
3	I	10	10	III	2086.7	6	Nonsingular	Stable
4	II	15(Annual)	13	I	2710.9	4	Nonsingular	Stable
5	II	15	13	III	2710.9	6	Nonsingular	Stable
6	II	20(Semiannual)	14	I	2819.96	4	Nonsingular	Unstable
7	II	20	14	III	2816.64	6	Nonsingular	Stable

(1) singularity of Hessian involves σ^2_w, σ^2_v, G.

Table 1. Summary of Identification Runs

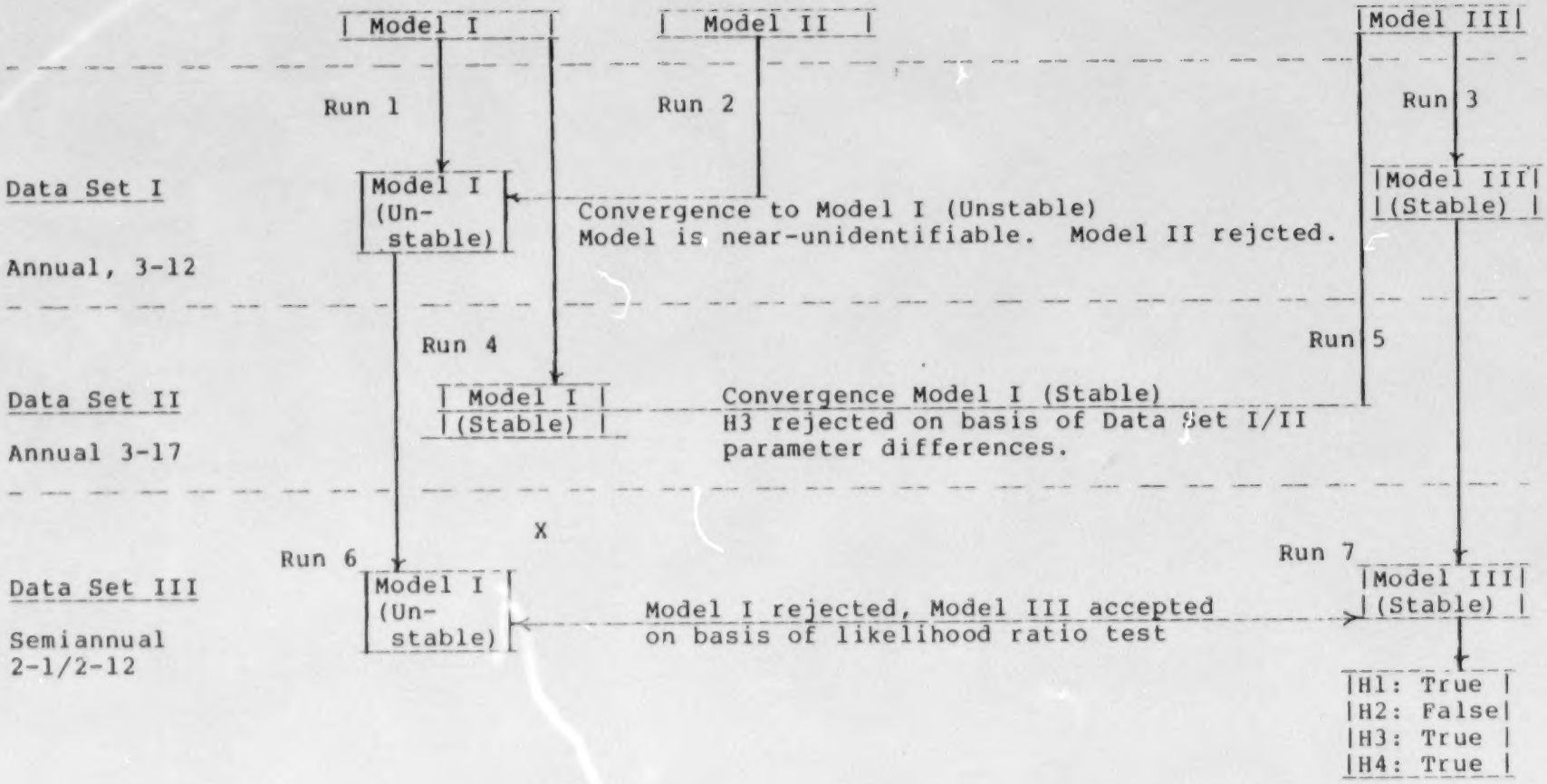


Figure 1

Identification Process

4.0/A/13--s.s.

Run	A	G	σ_w^2	σ_v^2	$V_{1\|0}$	D_R	σ_R^2
1	1.00(.01)	-	25.4(3.6)	20.7(2.6)	181.6(30.9)	-	-
2	.92(36)	1.25(3.5)	17.5(95)	23.7(35)	179.9(47) 152.8(26) 127.2(68)	-	-
3	.84(.01)	-	22.0(4.2)	20.6(3.1)	183.5(31.5)	28.1(12.9)	11.2(7.1)
4	.96(.01)	-	27.6(3.0)	18.2(2.3)	192.0(31.5)	-	-
5*	.95	-	28.1	18.0	192.4	0	0
6	1.00(.006)	-	12.56(1.52)	20.9(1.8)	168.0(28.4)	-	-
7	.914(.027)	-	11.0(1.8)	21.3(2.0)	170.2(29.0)	12.6(5.3)	3.2(1.7)

*Convergence incomplete.

Table 2. Parameter Estimates and Standard Deviations (-), Calculated from Hessian Matrix

closely agree with those made using Data Set I, (the time scale must be converted from full to half years). The difference in LLF's for models I and II is 2819.91-2816.64=3.32 so that the asymptotic likelihood ratio test criterion is χ^2=6.4, with two degrees of freedom. Thus Model I is rejected and Model III accepted. At the same time, H2 is rejected and H4 accepted.

Summary of Hypothesis Tests

H1, the hypothesis that the variance attributable to the stochastic input process is positive, is accepted, since σ^2_w>0 (statistically significant) in all models tested.

H2, the hypothesis of instability was rejected in favor of H4, the trend hypothesis. The superiority of Model III was striking when all data in the range 2-1/2 - 12 were included, but was not highly significant for Data Set I. Both models yield good fits to the data, and both predict significant changes in IQ over the age range 2-1/2 - 12.

H3, the hypothesis of constant parameters for the entire age range, is rejected. The unstable model of ages 2-1/2 - 12 is stabilized when ages 14, 15, 17 are included. This indicates that a model applying to all ages requires time varying parameters.

H4, the hypothesis of trend effects, is accepted on the basis of runs involving data set III. The hypothesis of positive correlations of initial IQ and trend is also accepted.

4.5 Model Validation

Whiteness of Residuals

Figure 1 exhibits sample values for correlations among lagged residuals for Model III and data set III. Notice that each sample value is smaller in magnitude than its corresponding standard error (except for one case), and that there is no discernible pattern to the values of the residuals. Comparison with Figure 2 indicates marginal superiority for Model III over Model I.

Residual Variances

Figure 3 exhibits predicted and observed values of residual variances by age. There is no pattern to suggest lack of model homogeneity by age or a drift away from the correct model by age.

Cross-Sectional Variance

Predicted values for total cross-sectional variance of the measurements are presented in Figure 4. The decrease in variance at ages 4 and 5 appears to us to be anomolous, possibly due to artifact in imputation of missing data. Otherwise the fit of real and calculated values appears to be good.

Simulation Results

As a test of Model III, 80 trajectories were simulated, plotted and assigned subjectively to "clusters". The results, plotted in Figure 5, should be compared to McCall's clusters in Figure 4.2.1, taking should take

Figure 1

Values of Lagged Correlations
Model III Data Set III

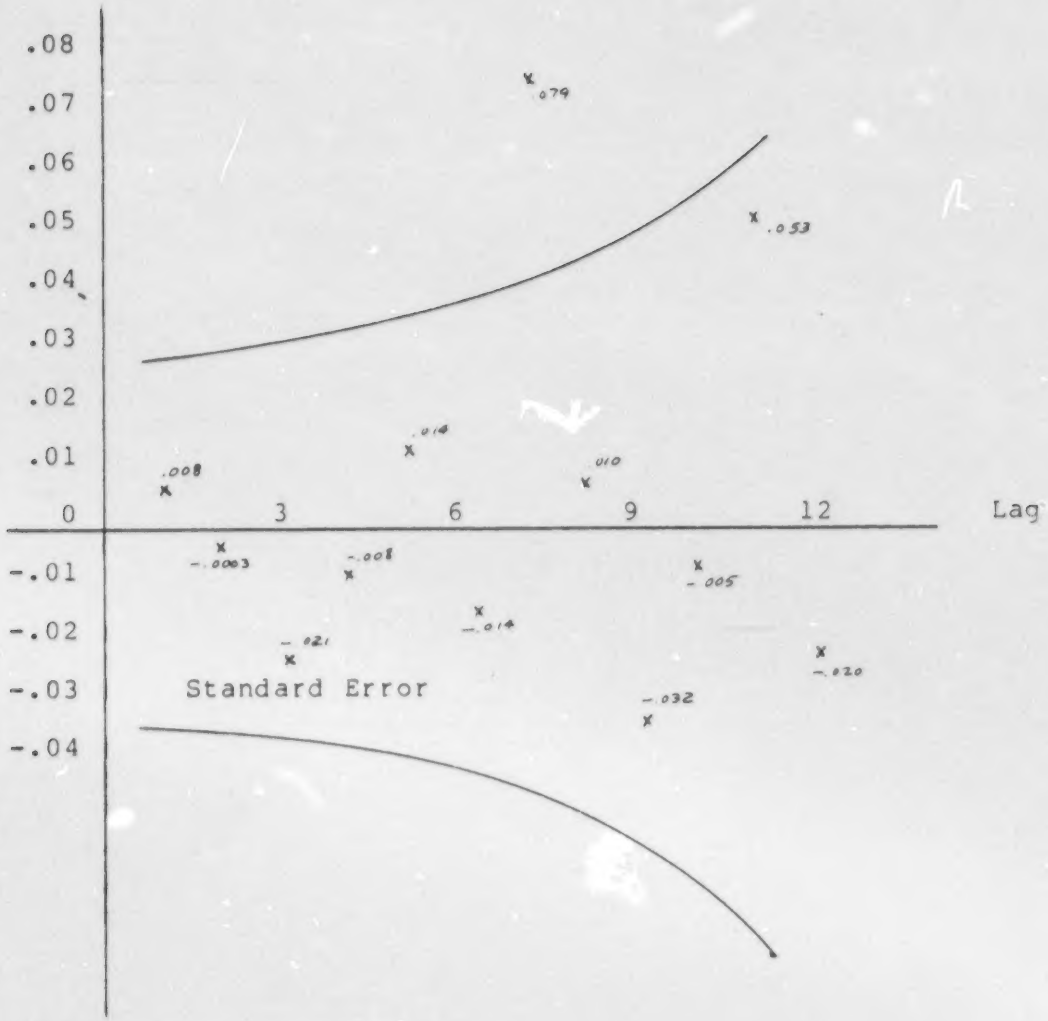
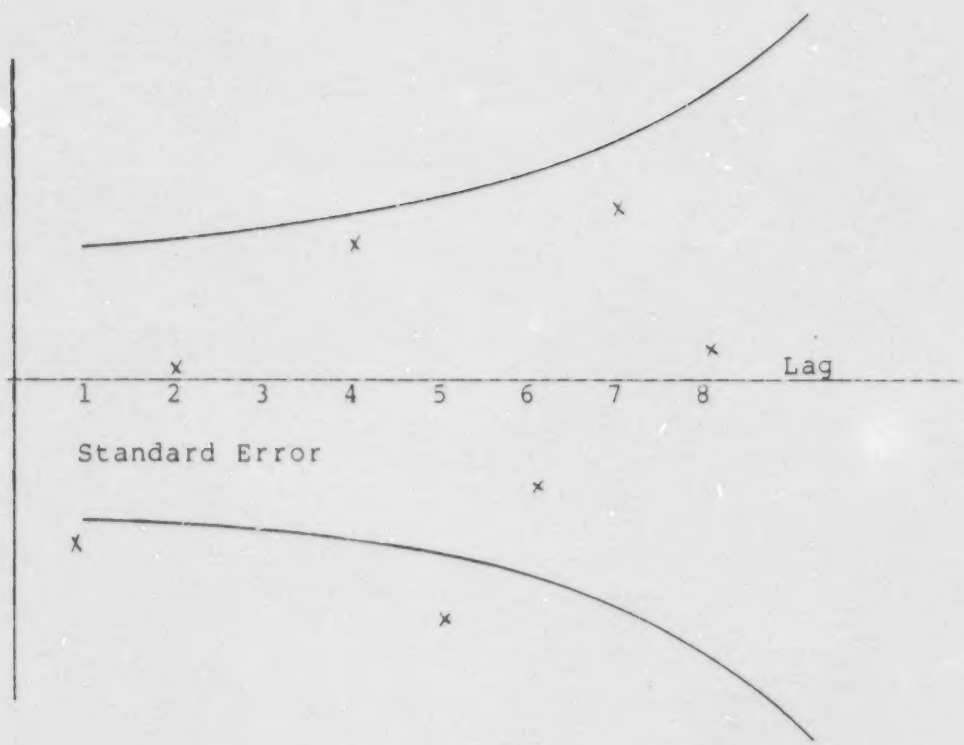


Figure 2

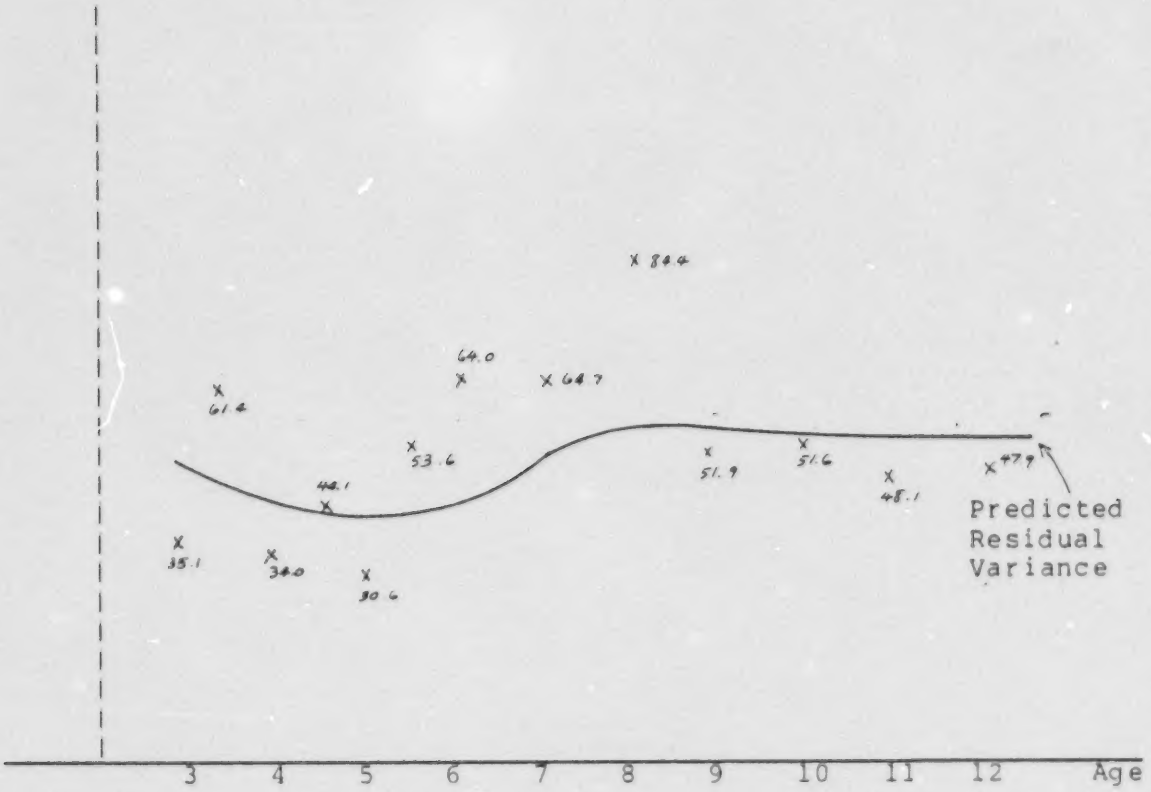
Values of Lagged Correlations
Model I Data Set I



Run 1. Sample Lagged Residual Correlations by Lag

Figure 3

Magnitudes of Residual Variances, Predicted and Observed



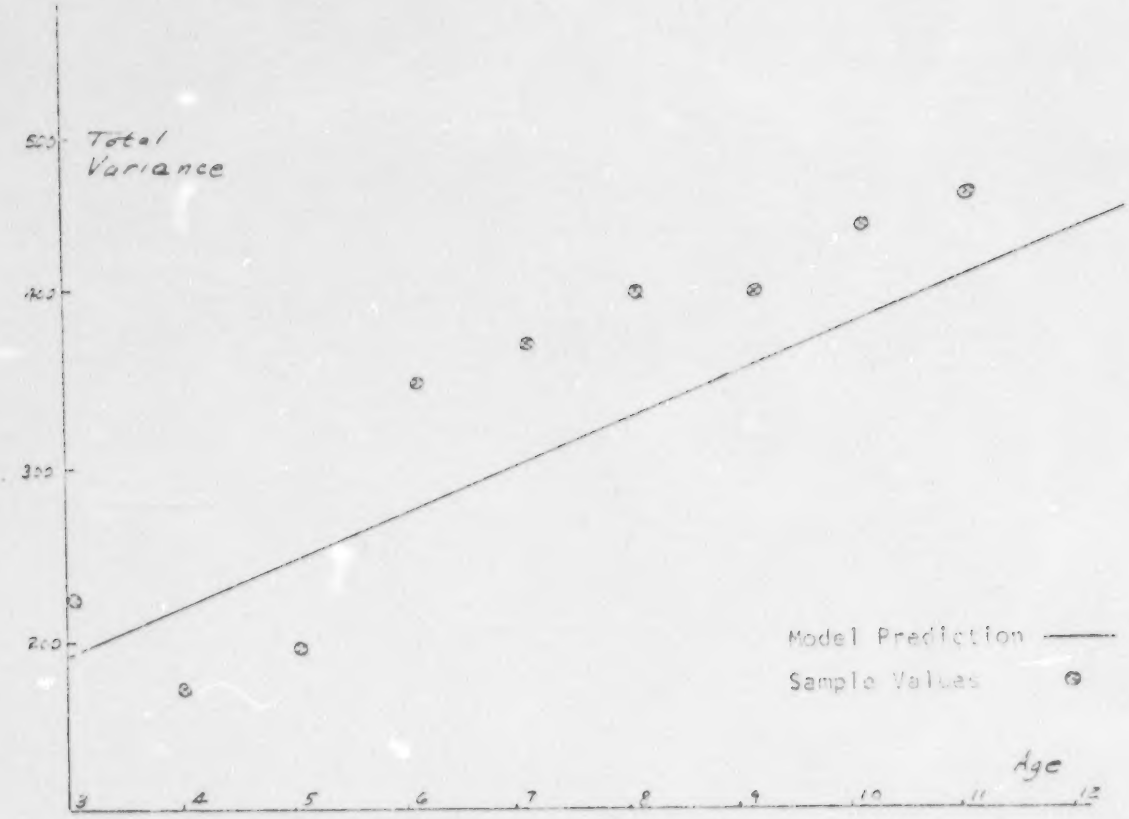
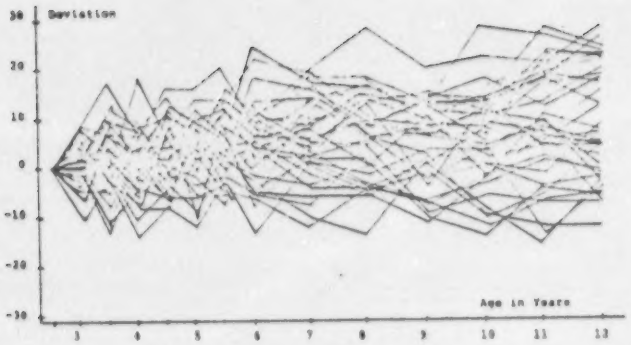


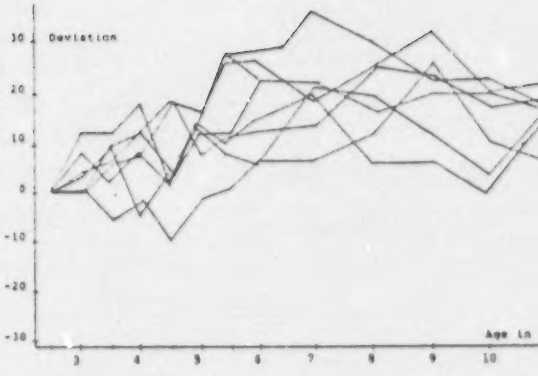
Figure 4. Model Prediction and Sample Values of Total Variance



Trajectory Group 1

36 Simulated Growth Curves

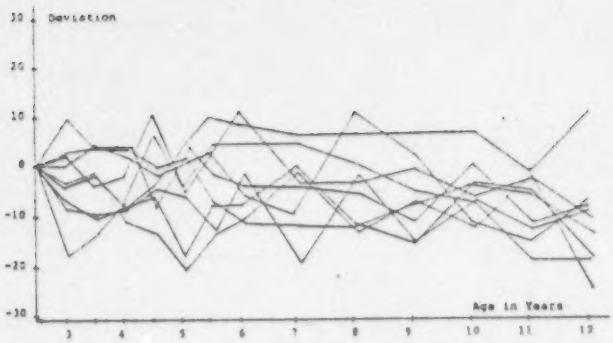
Figure 5A



Trajectory Group 4

6 Simulated Growth Curves

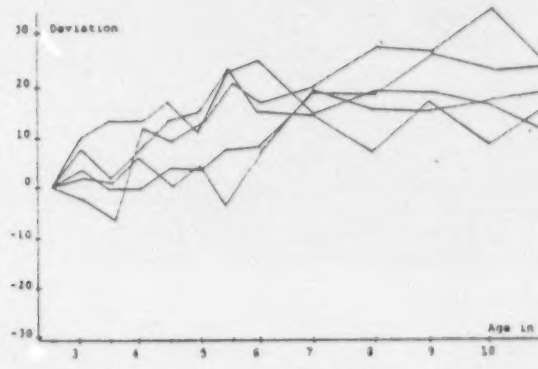
Figure 5D



Trajectory Group 2

9 Simulated Growth Curves

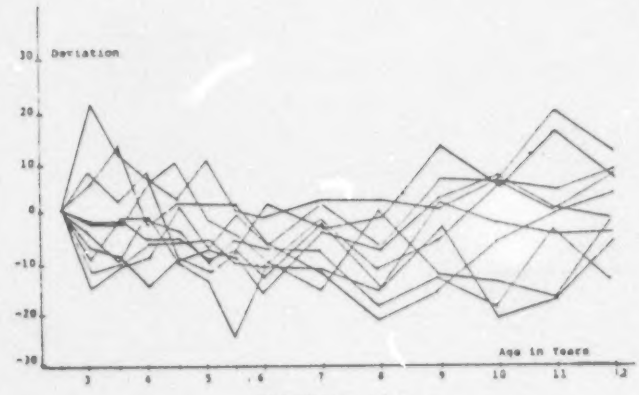
Figure 5B



Trajectory Group 5

5 Simulated Growth Curves

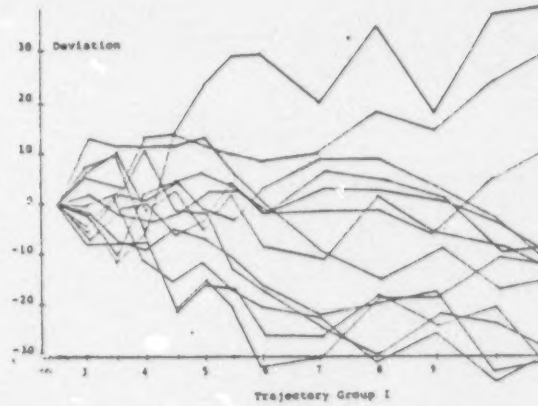
Figure 5E



Trajectory Group 3

10 Simulated Growth Curves

Figure 5C

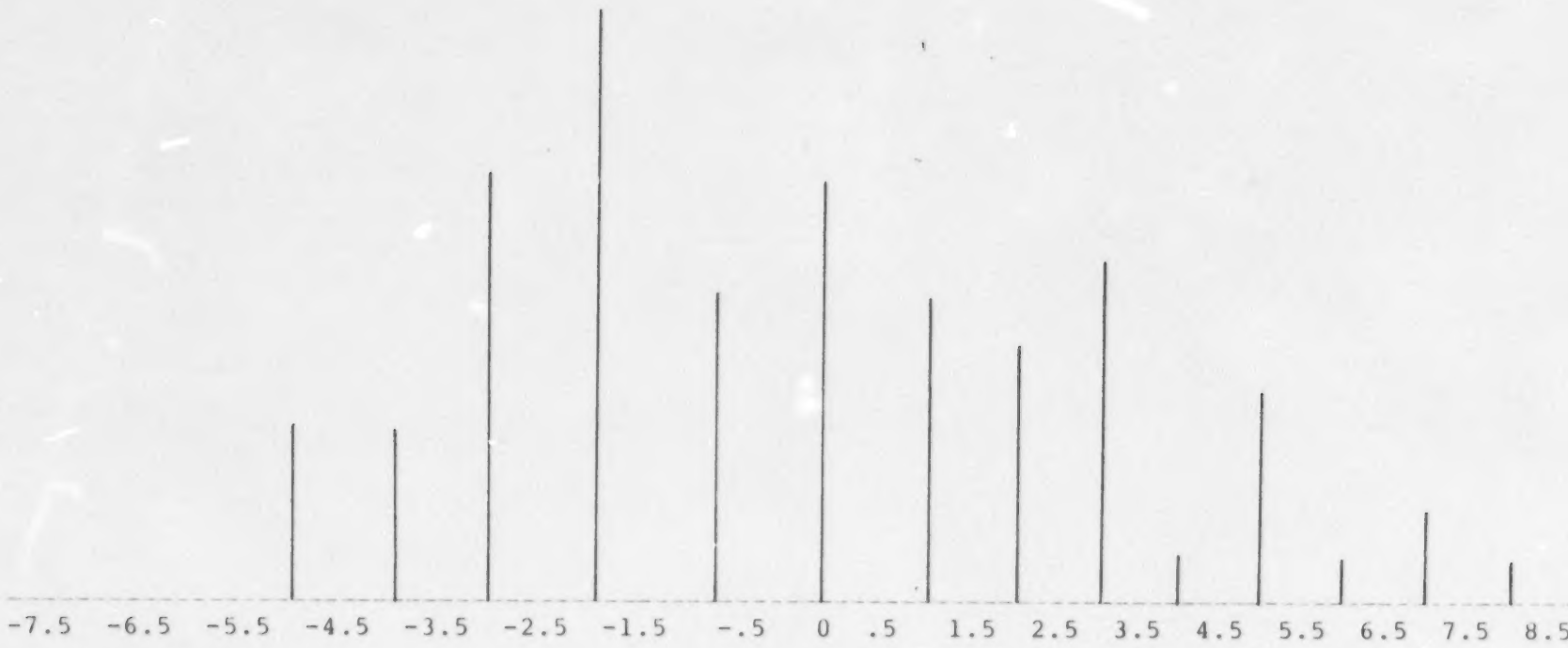


Trajectory Group 1

11 Simulated Growth Curves

Figure 5F

Figure 5. Eighty Simulated Trajectories Arranged Subjectively into Clusters



Distribution of Trend Term Estimates

Data Set I/Model III

Figure 6

account of three differences in plotting: (1) The simulated curves are plotted from a common origin, while McCall's are corrected for zero longitudinal mean, (2) The scale ratio in Figures 5A-F produces somewhat flatter curves than in Figure 4.2.1, and (3) the simulated curves are plotted only to age 12, the validity limit of the model, and so should be compared only to 2-1/2 - 12 year range of curves in Figure 4.2.1. The correspondence between these simulated groups of curves and the real clusters appears to us to be close, although this is a subjective judgment. The agreement makes plausible the ability of very simple stochastic models to "explain" relatively complex patterns of growth.

Distribution of Trend Levels

Figure 6 shows the distribution of estimates of trend levels R_i across cases i, obtained as Kalman filter state estimates at the conclusion of each trajectory. Goodness-of-fit to a normal distribution seems excellent, although there is apparent skewing.

4.6 Substantive Conclusions

Variance Due to Environment and Other Sources

A substantial part of total variance in IQ in the Fels population is apparently due to stochastic effects (environmental and/or human response) as well as to mean environmental characteristics. The latter source of variance has been long established, but quantitative investigation of stochastic effects is new. Based on the rather imprecise parameter estimates of Run 7 (Data Set III, Model III), the following decomposition of age 12 variances can be computed.

Source	Variance	Fraction
Initial State	139.7	.36*
Mean Environment	163.0	.42
Stochastic	64.7	.17
Measurement Error	21.3	.05**

Sources of IQ Variance at Age 12

* Upper bound for "heritability"
**Reliability = .95

These figures place "heritability" of IQ in the Fels population at .36 or less. Heritability figures quoted by Bloom range from .60 to .88, based on analyses that Layzer [46] has described as "numerology".

Trend Effects

The accumulated increment due to trend effects is (Model III) asymptotically equal to R/(1-A), where $^{2}_{R}$=11.2±7.1 and A=.84±.01. We calculate the difference between 10'th and 90'th percentile cumulative trends to be 13 to 26 points. In comparison, Bloom quotes estimates for cumulative effect of "extreme" environmental differences at about 20 points, citing Burks (1928) and Sontag (1958).

Stability

We have found that the growth model is only marginally stable below age 12. This leads to the conclusion that gains or losses in IQ that occur, even at very early ages, do not dissappear entirely within the 3-12 year time

frame. Furthermore, the process stabilizes at or about age 12 (H3). Thus effects of fluctuations may in fact never disappear. Our analyses do not address this question because of sparseness of longitudinal data after age 12.

Stabilization of IQ after initial periods of flux has often been noted. Charts in Bloom [8] from data collected by Bayley (1949) all indicate stabilization in the 10-12 year age range, in general agreement with our conclusion.

Correlations of IQ and IQ Change; the Overlap Hypothesis

Bloom has plotted correlations of $(IQ_{17}-IQ_x)$ with IQ_x, i.e., IQ change with IQ at age x, based on data from Bayley [84]. These generally negative correlations lead him to comment that he is "not clear as to why the correlations between initial scores and changes tend to be negative other than a general ceiling effect." Other researchers have noted such negative or very low correlations, which are, however, predicted by the present analysis.

Our model for the IQ at age k+1 given that at k is

$$X_{k+1} = AX_k + R + W_k,$$

with symbols as in Chapter II. This leads to the relation

$$Cov(X_{k+1}-X_k, X_k) = (A-1)VarX_k + Cov(R,X_k).$$

Since $A<0$, and $Cov(R,X_k)<0$, the IQ-IQ change correlation can be either positive or negative, depending on relative sizes of terms. Model III parameters predict a small <u>negative</u> correlation.

The hypothesis that $Cov(X_{k+1}-X_k, X_k)=0$ is known in developmental psychology as the "overlap hypothesis" and in engineering as "uncorrelated increments". Thus our Model I, where $\hat{A}=1$ and $Cov(R,X_k)=0$, leads to $Cov(X_{k+1}-X_k, X_k)=0$, corresponding to the overlap hypothesis. Under Model I, longitudinal correlations are explained by shared common evolution, i.e.,

$$Cov(X_{k+j}, X_k) = VarX_k.$$

For stable models, with $A<1$, the model predicts negative IQ-IQ change correlations, as are frequently observed.

Stanford-Binet Reliability

The runs give us an estimate of measurement error variance of approximately 20 points-squared, leading to a calculated test reliability* of .92 at age 6, increasing to .95 at age 12. These estimates are in agreement with generally accepted pychometrically based estimates although the method of calculation is altogether different.

Prediction of Scores at Maturity

Predictions of IQ scores at 18 years from sequences of earlier test data has been considered, e.g., by Bayley [84]. The problem can be solved in principle by identifying the growth system (as we have done) and using the associated Kalman filter for optimal predictions. We did not experiment quantitatively with this idea.

---

*"Reliability" is a psychometric term with the following definition, rooted in test theory. Let X be a "true" variable taking values in a population P and let Z be the measured value of X obtained from an instrument T. Then the reliability of T (for population P) is defined as VarX/VarT, where the variances are taken across P.

5.0/B/1

Chapter V Analysis of National Day Care Study Longitudinal Data

5.1 Introduction

This chapter discusses identification of a longitudinal data set from the National Day Care Study (NDCS) [89], a federally funded study of about 2000 three- and four-year-olds in a sample of day care centers in Atlanta, Detroit and Seattle. Our data set consists of 110 children who were given the Preschool Inventory Test (PSI) [49] at each of 4 data collection "points" that for practical purposes can be considered as equally spaced at 6 month intervals. The PSI is an inventory of skills, verbal abilities and knowledge, highly correlated with Stanford-Binet IQ scores. The 32 item version (selected from 64 original items) was used. Psychometric details are found in Interim Report 2, Volume II, National Day Care Study (1976).

It is not the purpose of this chapter to form conclusions about the effects of day care on cognitive growth. That is the subject of the NDCS itself, to be discussed in the NDCS final report to be published in Summer, 1978. In this chapter we treat data merely as a problem in structure determination and parameter identification, using methods developed in the thesis. The results of this identification will be used in the NDCS analysis proper.

The data consist of 110 sequences of 4 points each; for such short sequences identification theorems 3.6.2 and 3.6.3 are of practical importance and lead to the results quoted in the following table, depending on equilibrium (steady state) or disequilibrium of the system. We expect growth systems to be in <u>disequilibrium</u> initially, but we must expect less robust estimates when the sequence length is not sufficient for identifiability of the steady state system.

Model Order r	Disequilibrium* $\sigma_R^2 = 0$	Equilibrium** $\sigma_R^2 = 0$	Equilibrium** $\sigma_R^2 > 0$
1	3	3	4
2	4	5	6
3	5	7	8

Minimum Sequence Lengths for Identifiability

*Here we assume "full rank disequilibrium", as defined in Section 3.6.

**We assume that the controllability condition of Theorems 3.6 and 3.7 is satisfied, as is the case for the steady state (stationary) system.

From this we learn that models of order greater than r=2 are not identifiable(4), and that identification of the second order model is restricted. Three models can be considered. We present them in full form (not corrected for zero mean), but partition the parameter $\theta = (\theta_I, \theta_{II})$ as described in Section 3.4.

Model I. First Order without trend input.
The specific model is

$$X_{k+1} = AX_k + W_k + U_k$$

$$Z_k = X_k + V_k,$$

with parameter vectors

$$\theta_I = (A, \sigma_w^2, \sigma_v^2, V_{1|0})$$

$$\theta_{II} = (\bar{X}_{1|0}, U_1, U_2, U_3)$$

Model II. Second order without trend input.
The specific model is

$$X_{k+1} = \begin{bmatrix} A_1 & A_2 \\ 1 & 0 \end{bmatrix} X_k + \begin{bmatrix} G_2 \\ 1 \end{bmatrix} W_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_k$$

$$Z_k = (0\ 1)\ X_k + V_k,$$

with parameter vector

$$\theta_I = (A_1, A_2, G_2, \sigma_w^2, \sigma_v^2, \underline{V}_{11}, \underline{V}_{12}, \underline{V}_{22})$$

$$\theta_{II} = (X_{1|0}, U_1, U_2)$$

where the parameters $\underline{V}_{11}, \underline{V}_{12}, \underline{V}_{22}$ are components of $V_{1|0}$ and $X_{1|0}$ has 2 components. Note that this model is identifiable under the assumption of full fank disequilibrium, but not in general. For instance, the stationary model is not identifiable since the Toeplitz covariance matrix of the sequences has only 4 independent elements available to identify 6 elements of θ_I.

Model III. First Order with trend input.
The specific model is

$$X_{k+1} = AX_k + W_k + R + U_k$$

$$Z_k = X_k + V_k$$

with parameter vector

$$\theta_I = (A, \sigma_w^2, \sigma_v^2, V_{1|0}, D_R, \sigma_R^2)$$

$$\theta_{II} = (\bar{X}_{1|0}, U_1, U_2, U_3)$$

The sufficient statistics of the problem include four cross-sectional means and 10 independent elements of the covariance matrix (under full rank disequilibrium). The steady state covariance matrix has only four in dependent elements. Threefore, under stationarity, only Model I is identifiable. Identification of Models II, III must rely on displacement of the growth system from equilibrium. Further details about these models were presented in Chapter II. The partitioning of parameters $\theta = (\theta_I, \theta_{II})$ was discussed in Section 3.4; we wil identify the model using the stepwise procedure outlined there, beginning wit identification of model order r and θ_I based on the data corrected for zero mean.

5.2 Identification of Fluctuation Model

Tables 1, 2, and 3 summarize results of three series of runs (one for each model), based on data corrected for zero mean. The following paragraphs describe the process leading to determination of structure to be Model I.

(1) The likelihood ratio criterion gives ambiguous results. At the level of significance .05, Model I is accepted and others rejected, but we would accept Model II at the level .08. The Akaike AIC criterion would result in the choice of Model III.

(2) Only Model I was associated with a Hessian well enough conditioned to use in the minimization algorithm. Partly as a result, only Model I was identifiable in the sense of providing meaningfully precise parameter estimates (see Table 2). In Table 3 it is shown that estimates for G, σ^2_w, σ^2_v are correlated nearly perfectly so that they are "confounded" in the statistical sense.

(The matrices in Table 3 are prepared by converting the inverse information (Hessian) matrix into the form of a correlation matrix by dividing each off-diagonal entry by the SD's associatd with the appropriate rank and column. Diagonal elements are replaced by their square roots, equal to Cramer-Rao lower bounds for parameter estimate variances. Singularity of the Hessian appears as very large diagonal entries and as off diagonal entries with absolute values near one. Since the Hessian matrix represents the curvature of the LLF surface, it is an indication of the robustness of parameter estimates.)

Model	Conditioning of Hessian	LLF	θ_I, Number of Parameters***	Level of Significance
Theoretical Lower Bound*	-	794.32	10	.14
Model I	Good	799.20	4	Null Hypothesis
Model II	Poor	795.42**	8	.10
Model III	Poor	796.7	6	.08
Model III/ Local Minimum	Poor	799.16	6	.96

*See Section 3.7 for a description of this bound.
**Upper Bound. Full minimum not achieved.
***The number of parameters in the covariance matrix, assuming stationarity is 4. Assuming full rank disequilibrium, there are 10.

Summary of Statistical Test Data

Table 1

5.0/B/6--s.s.--12 ptch spacing

	$A^{\cdot}$	G_2	σ_w^2	σ_v^2	$V_{1\|0}$	D_R	σ_R^2
Model I	.909(.05)*	-	3.95(1.52)	5.18(1.10)	17.38(3.07)	-	-
Model II	1.093(.71)	.392(.63)	10.83(22.7)	0(19.0)	18.92(29.2)	-	-
	-.194(.67)				15.04(2.5)		
					14.78(5.0)		
Model III	.347(.80)	-	9.96(13.1)	.06(14.2)	20.1(14.5)	8.6(11.8)	6.1(17.6)

*Standard deviations from Hessian matrix in parentheses.

Parameter Estimates for Three Models

Table 2

102

103

5.0/B/7--s.s.—12 pitch spacing—1.m.8

	A_1	σ_w^2	σ_v^2	$V_{1\|0}$
A_1	.519*	-.747	.673	-.399
σ_w^2	-.747	1.522	-.835	.307
σ_v^2	.673	-.835	1.099	-.329
$V_{1\|0}$	-.399	.307	-.329	3.074

Model I. First Order/No Trends.

	A_1	A_2	G_2	σ_w^2	σ_v^2	$V_{1\|0}$ Components (1,1)	(1,2)	(2,2)
A_1	.714	-.991	.629	-.879	.838	-.830	-.203	.641
A_2	-.991	.671	-.711	.926	-.894	.886	.184	-.660
G_2	.629	-.711	.627	-.888	.931	-.923	-.123	.360
σ_w^2	-.879	.926	-.888	22.741	-.994	.985	.178	-.632
σ_v^2	.838	-.894	.931	-.994	18.998	-.991	-.170	.581
$V_{1\|0}$								
(1,1)	-.830	.886	-.923	.985	-.991	-9.167	.276	-.533
(1,2)	-.203	.184	-.123	.178	-.170	.276	2.503	.380
(1,3)	.641	-.660	.360	-.632	.581	-.533	.380	5.001

Model II. Second Order/No Trends.

	A_1	σ_w^2	σ_v^2	$V_{1\|0}$	D	σ_R^2
A_1	.807	-.966	.985	-.968	-.991	-.997
σ_w^2	-.966	13.124	-.994	.977	.952	.955
σ_v^2	.985	-.994	14.209	-.982	-.973	-.977
$V_{1\|0}$	-.968	.977	-.982	14.466	.974	.965
D_R	-.991	.942	-.973	.974	11.805	.996
σ_R^2	-.997	.955	-.977	.965	.996	17.554

Model III. First Order/Trends

Table 3. Correlations and Standard Deviations from Hessian Matrix

*Standard Deviations are displayed on diagonal, correlations on off-diagonal.

(3) For Models II and III the parameter estimate for measurement error is zero or near zero, although this figure is known from independent psychometric analyses to lie in the range 4-6.

(4) A second local minimum was found for Model III, with parameter estimates and LLF value essentially equal to those found for Model I. (Models of form I constitute the subspace $\{\theta | D_R = \sigma^2_R = 0\}$ of the space generated by Model III parameters.) The Model III parameter estimates quoted in Table 2 refer to the supposed global minmum LLF and are radically different from the Model I solution. The estimate A=.347 (global minimum) gives a poor fit to the mean growth data. Figure 5.4.1 shows that the sample mean growth curve is nearly linear, while the value A=.347 would lead to extreme convexity given nearly constant deterministic input sequence $\{U_k\}$. Thus the Model III parameters do not yield a realistic growth model.

Validation of Model I: Absolute Goodness of Fit

The following tests of goodness-of-fit were performed to validate selection of Model I.

Whiteness of Innovations

Table 4 displays sample correlations (r) of innovations at lags 1,2, and 3. Under the null hypothesis these correlations are sample values from a bivariate normal distribution with covariance zero, and have the t-distribution in the transformed variable

$$r\sqrt{\frac{n-2}{1-r^2}}$$

where n is the number of cases used. None of the t-values so calculated is significant in a two-sided test, even for levels of significance as conservative as .50. Thus the hypothesis of whiteness is accepted.

Lag	Sample Correlation	Innovations Included	Number of Cases	t	DF
1	.040	T1-T2, T2-T3, T3-T4	240		238
2	.047	T1-T3, T2-T4	160	.59	158
3	.019	T1-T4	80	.17	78

Table 4. Innovations Correlation Data

Covariance of Measurement Sequences

The parameter estimate $\theta_I = (A, \sigma^2_w, \sigma^2_v, V_{1|0})$ can be used to estimate the covariance matrix among measurements at the four times (T1,T2,T3,T4) using the formula

$$\hat{\Sigma} = \begin{bmatrix} A^3 \\ A^2 \\ A \\ I \end{bmatrix} V_{1|0} \begin{bmatrix} A^3 & A^2 & A & I \end{bmatrix} + \begin{bmatrix} A^2 & A & 1 & 0 \\ A & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sigma^2_w + \sigma^2_v I,$$

derived in Appendix 4.
Calculation yields

$$\hat{\Sigma} = \begin{bmatrix} 24.89 & 17.34 & 15.12 & 13.05 \\ 17.34 & 24.26 & 16.64 & 14.36 \\ 15.12 & 16.64 & 23.49 & 15.80 \\ 13.05 & 14.36 & 15.80 & 22.56 \end{bmatrix},$$

while the sample covariance matrix is given by

$$S = \begin{bmatrix} 22.76 & 16.59 & 16.04 & 12.58 \\ 16.59 & 26.08 & 20.51 & 14.24 \\ 16.04 & 20.51 & 28.98 & 17.03 \\ 12.58 & 14.24 & 17.03 & 20.18 \end{bmatrix},$$

We wish to test the hypothesis $H_o: S=\hat{\Sigma}$. An approximate (likelihood ratio) test is given by Anderson [55], Section 10.8:

$$-2 \log \quad = N[T\text{-trace} \quad S^{-1} + \log|S^{-1}|]$$

$$-2 \log \quad \chi^2_{T(T+1)/2}$$

Calculation yields $\chi^2_{10}=9.65$, corresponding to the fractile value .53 in the distribution of χ^2. Thus H_o is accepted.

Robust Estimates

Although ML estimates are asymptotically efficient for normally distributed data, they may not be robust with respect to severe departures from normality, especially when sample sizes are small. Recently much attention has been paid to so-called robust statistics, especially the jackknife method developed by Tukey [86] and explained in detail by Mosteller and Tukey [85]. The data are divided into groups and the calculation (in this case an ML estimation) is made for each reduced body of data obtained by excluding one of the groups, as well as for the entire data set. Then, "pseudo-values" Y^*_j are defined by

$$Y^*_j = KY_o - (K-1)Y_j, \qquad j=1,2,..,k,$$

where k is the number of groups, Y_o the estimate for the entire data set and Y_j that obtained for the estimate omitting the j'th group. Then the jackknife estimate and it variance are given by

$$Y^* = \frac{1}{k} \sum_j Y^*_j$$

$$S^{*2} = \frac{1}{k} \varepsilon^2,$$

where S^2 is the sample variance of the pseudo-values.

We formed 11 groups of 10 subjects each and repeated the ML estimation for each of the 11 resulting reduced data sets, obtaining the results exhibited in Table 4. Each numerical minimization was carried out until the norm of the gradient was .000005 or less. It can be seen that both overall ML estimates and information matrix parameter error estimates are close to jackknifed values.

ML Estimates	A	σ_w^2	σ_v^2	$V_{1\|0}$
Entire Sample	.90949	3.94711	5.17759	17.37589
1	.90586	3.57428	5.14253	18.16757
2	.90547	4.56481	4.77656	19.00743
3	.90223	3.78566	5.31813	17.90313
4	.90802	4.22483	5.05623	16.90015
5	.91361	4.11079	5.07464	17.33767
6	.90841	4.07115	5.15885	17.03517
7	.92226	3.00405	5.74061	17.40330
8	.89182	5.28138	4.53196	16.43901
9	.90115	4.54255	5.05568	16.89951
10	.90642	3.73532	5.33622	17.14301
11	.93134	2.69613	5.59384	16.40833
Jackknife Estimate	.91249	3.79007	5.28226	17.8218
SD of Jackknife Estimate	.0348	2.414	1.035	2.332
SD from Information Matrix	.0519	1.522	1.099	3.071

Table 4. Estimates Obtained by Jackknife Method

5.3 Estimation of Mean Growth Curve

Now we accept $\theta_I=(A,\sigma^2_w,\sigma^2_v,V_{1|0})$ and Model I structure as given and estimate $\theta_{II}=(X_{1|0},U_1,U_2,U_3)$ using the methods of Appendix 6. In this way we obtain an estimate of the mean growth curve. The following is a summary of results. The mean growth model, discussed in Section 3.4, is

$$\begin{aligned}
\bar{X}_{k+1} &= A\bar{X}_k + U_k + \tilde{w}_k \\
\bar{Z}_k &= \bar{X}_k + \tilde{v}_k \\
\bar{X}_1 &\sim N(\bar{X}_{1|0}, V_{1|0}/N) \\
\tilde{w}_k &\sim N(0, \sigma^2_w/N) \\
\tilde{v}_k &\sim N(0, \sigma^2_v/N)
\end{aligned} \tag{1}$$

where the following parameter values were obtained by the method discussed in the previous section.

$$\begin{aligned}
A &= .909 \\
V_{1|0} &= 17.38 \\
\sigma^2_w &= 3.95 \\
\sigma^2_v &= 5.18
\end{aligned} \tag{2}$$

Recall that the input sequences $\{U_k\}$ drives the system into mean growth, as displayed in Figure 5.4.1. Under the hypothesis that the input U_k is <u>constant</u>, i.e., $U_1=U_2=U_3$, we obtain the following estimates.

$$\bar{X}_{1|0} = 11.30$$
$$U_1 = U_2 = U_3 = 4.89 \qquad (3)$$

These values can be used to compute the following predicted mean growth curve points, which are compared with sample means. These data are graphically presented in Figure 1.

	T1	T2	T3	T4
Sample Mean	11.37	14.89	18.16	21.96
Predicted Mean (based on (3))	11.30	15.16	18.67	21.87

Using the technique presented in Appendix 6, we can test goodness of fit under the null hypothesis that parameter estimates given in (2) and (3) are correct and that the model structure has been correctly determined. The resulting chi-square test is

$$\chi^2 = 4.10$$
$$DF = 2$$
$$P = .13$$

Thus we accept the null hypothesis at the level P=.13, including the hypothesis $U_1=U_2=U_3$. Since the goodness of fit criterion does not warrant relaxation of the constraint $U_1=U_2=U_3$ at the level P=.13, we will not consider more complex cases involving nonconstant deterministic input.

Figure 1 illustrates twenty randomly selected data sequences, together with the sample mean trajectory and fitted mean points. Although the sample mean growth curve is fitted accurately, small departures at k=2,3 indicate that the true mean may evolve slightly more nearly linearly than predicted by the estimate of A(.909) based on fluctuation data.

5.4 Construction of Generalized Change Scores

The sample of day care centers in the National Day Care Study was selected to include a wide range of staff-child ratios, teacher qualifications, etc. In order to analyze the effects of these variations, we require an estimate of the growth rate experienced by each subject. The Generalized Change Score (GCS) is based on the model in ARMA (rather than state space) form:

$$\boxed{\begin{aligned} A(b)Y_{k+1,i} &= B(b)W_{ki} + E_{ki} \\ Z_{ki} &= Y_{ki} + V_{ki} \end{aligned}} \qquad (1)$$

We have temporarily reintroduced indexing by subject (i) and have subsumed all inputs in the single term E_{ki}. The polynomials A(b) and B(b) are considered to have been identified by procedures that have been discussed. Equations (1) imply that

$$A(b)Z_{k+1} = A(b)V_{k+1} + B(b)W_k + E_k, \qquad (2)$$

an equation that motivates the following.

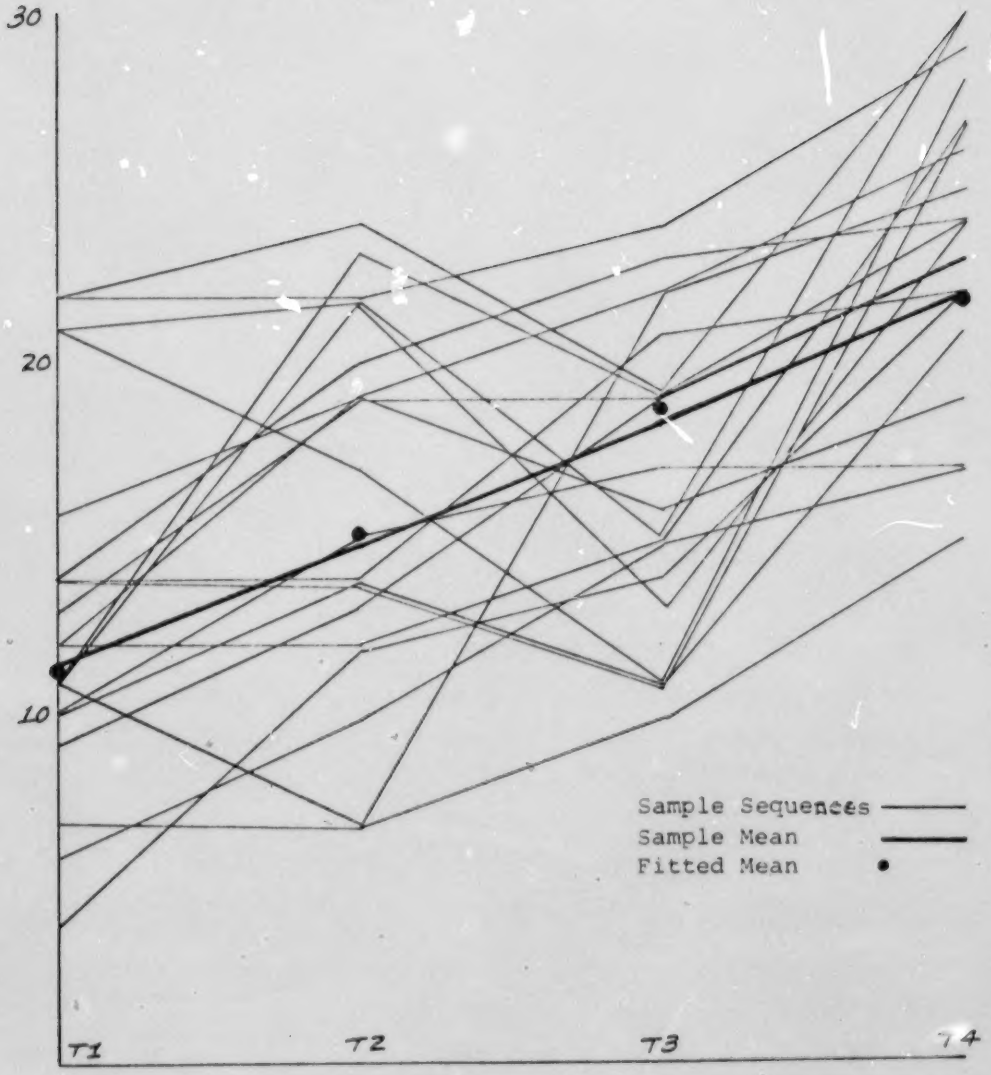


Figure 1. Sample Mean and Randomly Selected Trajectories

Definition Let $\{Z_k\}$ satisfy the model (1). Then the GCS of Z_{k+1} is given by the expression

$$\Delta_{k+1} = A(b)Z_{k+1} = Z_{k+1} - A_1 Z_k - \ldots - Z_p Z_{k+1-p} \quad (3)$$

Theorem The GCS is an unbiased estimator of the total input E_k.

Proof: The mean and variance of Δ_{k+1} are, from (2),

$$E(\Delta_{k+1}) = E_k \quad (4)$$

and $$\text{Var}(\Delta_{k+1}) = (1+A_1^2+\ldots+A_r^2)\sigma_v^2 + (1+B_1^2+\ldots+B_{r-1}^2)\sigma_w^2 + \sigma_{E_k}^2 \quad (5)$$

Thus Δ_{k+1} is an unbiased estimator for the total input effect for each subject.

Note that no claim is made for efficiency of the GCS as an estimator for E_k. More efficient estimators can be constructed using the innovation

$$I_{k+1} = Z_{k+1} - H\hat{X}_{k+1|k},$$

based on the Kalman filter forecast of Z_{k+1}. Although this technique is a promising one, it is not discussed further in this thesis because of the complexity of the approach compared to our GCS.

In the National Day Care Study, where the data have been identified as Model I (first order AR), the GCS calculated on the basis of parameter estimates discussed here is

$$\Delta_{k+1} = Z_{k+1} - AZ_k, \quad (6)$$

with $\hat{A}$=.909. Equation (5) can be used to estimate an error variance of the GCS of approximately 12.4. GCS scores were calculated and attached to each child in the larger sample of children who were given PSI tests at T3 and T4, and these scores regressed on a variety of different possible predictors of child outcome, such as average group size, numbers of teachers, etc. These essentially static analyses were conventional, and detailed description would go beyond the intended scope of this thesis, which is limited to dynamic aspects of the problem.

5.5 Summary and Conclusions

The computations reported in this chapter verify that robust identification of ARMA models can be carried out even with very short sequences, provided that conditions of the identifiability theorems are satisfied. Moreover, identification of the system leads to construction of a generalized change score that has proved useful in the National Day Care Study. We do not come to important substantive conclusions about the PSI test or about day care effects on child development however. We refer the interested reader to reports and publications of the National Day Care Study [89].

6.0/A/1

Chapter VI Summary and General Conclusions

6.1 Summary of Results

This thesis has developed a methodology for estimating and identifying models that describe short non-stationary time series from cross-sectional data. This extends the applicability of time series methods from the traditional areas, primarily in engineering and economics, to growth curve analyses and other potential applications where series are inherently short and/or nonstationary. Our technique uses the Kalman filter recursive equations, and their gradients with respect to the parameter vector, to compute the exact likelihood function and its derivatives. The Newton-Raphson and gradient methods are used to obtain ML parameter estimates for specific model structures. Structure determination is carried out by use of the likelihood ratio criterion and by examination of the Hessian matrix for evidence of non-identifiability. A great deal of practical experience in applying the methodology was acquired, in hundreds of separate computer runs.

The methodology was applied to two data sets involving measures of cognitive development. In both cases, noisy first order autoregressive models achieved close statistical fits to the data, measured by the criteria of (1) whiteness of residuals and (2) goodness-of-fit to the covariance matrix of measurement sequences. Models of the type hypothesized are new in developmental psychology and are of theoretical interest because they separate measurement error from process variance induced by environmental fluctuations, and because they provide valid statistical tests for hypotheses about stability, trend effects, and other system properties. Moreover, identification of the growth system provides a basis for construction of change measures for application in educational evaluations.

The thesis also presents new theoretical results. Theorems are presented that state the conditions for global and local identifiability of short non-stationary cross-sectional time series, and that relate identifiability and rank of the Hessian matrix.

6.2 Useful Extensions

(1) The most useful extension may be to the problem in which short time series segments are collected from a nonstationary process at different levels of development (i.e., age of the nonstationary system). For example one might collect cognitive development data at different starting ages, so that the problem is to piece together the entire development of the system from segments of data, typified in the data set displayed in Figure 1. This objective seems feasible as an extension of the current methods. Essentially, one pieces together the developments of the state covariance at a particular age as the a priori state distribution for series beginning at that age. The evolution of the covariance is known as a function of system parameters. The likelihood function for a particular data set is evaluated by generating the implied state covariance matrix as a function of age and applying the Kalman filter algorithm just as is done in the single age case.

(2) In the model of Chapter II deterministic forces $\{U_k\}$ are assumed to be parameters to be estimated. One may change this formulation, assuming that $\{U_k\}$ is known, but that its coupling to the system is not. This is a formulation closer to conventional transfer function models. It may be of value when one wishes to evaluate the influence of known inputs on the developmental system, as in educational evaluation. In particular, measurements of family interactions could be related to cognitive development.

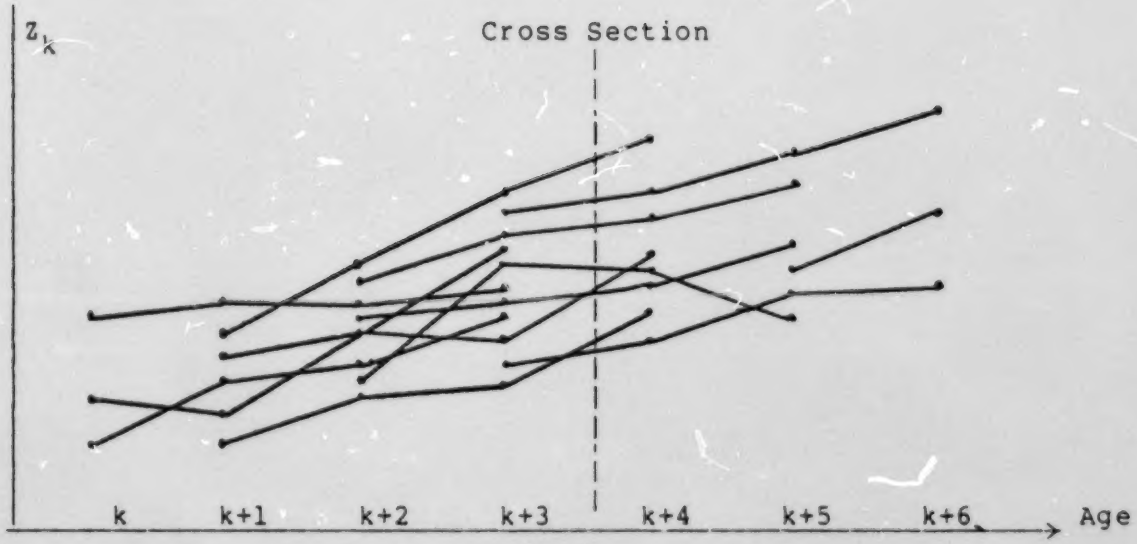


Figure 1

Ensemble of Short Time Series
at Different Starting Ages
(Multiple Cohorts)

(3) In applying the filter to the data, Bayesian estimates are obtained for trend terms R, case-by-case. For very short series these estimates are of little value. However, the estimates approach true values as series length approaches ∞. In series of moderate length the estimates for R may allow estimation of individual trajectories that are closely adjusted to individual growth rates. The value of such estimates for evaluating treatment effects is obvious. For instance it is possible that sequences of medical measurements of patients could be used to evaluate or forecast patient condition for use in diagnosis.

6.3 Other Applications

The two applications treated in this thesis involve cognitive growth data, and in fact the method was originally conceived to handle such data. But there are other applications that share the problems of these data: short length of individual series, and nonstationary nature. Among the possibilities are:

(1) Ecological Series Ecological series are likely to be nonstationary because of external input forces and in some cases are not stable. In many cases large cross-sectional data sets can be assembled, for instance by making measurements in multiple locations. Such data bases are identifiable using the methods of this thesis. The results can be useful both towards understanding ecosystem dynamics, particularly stability, and in diagnoses of the state and trend of particular systems.

(2) Non-stationary economic or sociological series, in which, e.g., 50 different states provide samples of 10 annual points of some parallel measure, e.g., prison populations. The cross-sectional method allows us to build statistical power without resorting to the assumption of stationarity implied when series are extended over several decades and hence over different historical contexts.

Appendices

APPENDIX 1

Calculation of Gradient and Hessian Matrix

The parameter estimation technique used in the computer program involves minimization by numerical methods of the scaled log likelihood function

$$L = \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{T} (I^2_{ki}/P_k + \log P_k). \qquad (1)$$

where I_{ki} and P_k are the innovation and its variance, respectively, as defined in Section 3.2. Minimization is carried out using the gradient, Newton-Raphson, or Davidon-Fletcher-Powell algorithm. All three methods require computation of the gradient of L, while the N-R technique also requires an approximation to the Hessian matrix. The Hessian matrix is useful in any case as an aid in system identification and in computing Cramer-Rao bounds to parameter estimate error variances. This appendix presents computations of the gradient and an approximation to the Hessian.

The gradient of (1) can be calculated to be

$$\underline{L} = \sum_{i=1}^{N} \sum_{k=1}^{T} \left\{ -I_k H \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} - \frac{1}{2} I^2_k P^{-1}_k \frac{\partial P_k}{\partial \theta} + \frac{1}{2} \frac{\partial P_k}{\partial \theta} \right\} P^{-1}_k, \qquad (2)$$

where P_k, I_k are scalars, $\hat{X}_{k|k-1}$ is sx1, and θ is mx1. For convenience, the case index i has been suppressed in the summand. This expression must be calculated exactly (except for round-off error) in order to locate local extrema exactly. The derivative of the state prediction $\hat{X}_{k|k-1}$ and of the innovations variance P_k, which occur in (2) are calculated as computational intermediates, using equations presented in Table 3.2.1.

The derivative $\frac{\partial \hat{X}_{k|k-1}}{\partial \theta}$ is obtained by differentiating the state prediction equations (3.2.2a) and (3.2.2b) of the filter algorithm, yielding

$$\frac{\partial \hat{X}_{k+1|k}}{\partial \theta} = F(I-K_k H)\frac{\partial \hat{X}_{k|k-1}}{\partial \theta} + \frac{\partial F}{\partial \theta}(\hat{X}_{k|k-1} + K_k I_k) + F\frac{\partial K_k}{\partial \theta} I_k. \quad (3)$$

Equation (3) gives a recursive calculation for $\frac{\partial \hat{X}_{k+1|k}}{\partial \theta}$ in terms of the state prediction $\hat{X}_{k|k-1}$ and the innovation I_k. Calculationally, the recursion for the state prediction and derivatives are carried out together by annexing (3) to the conventional Kalman equations. The filter gain gradient equation and variance gradient equations (Table 3.3.1) are obtained by straightforward differentiation of (3.2.2).

The Hessian matrix is obtained by differentiating the column vector given in (2) to form an m x m matrix:

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \sum_{i=1}^{N}\sum_{k=1}^{T}\left\{-I_k H\left(\frac{\partial^2 \hat{X}_{k|k-1}}{\partial \theta_i \partial \theta_j}\right) + H\left(\frac{\partial \hat{X}_{k|k-1}}{\partial \theta_i}\right)\left(\frac{\partial \hat{X}_{k|k-1}}{\partial \theta_j}\right)' H'\right.$$

$$+ I_k H\left(\frac{\partial \hat{X}_{k|k-1}}{\partial \theta_i}\right)\frac{\partial P_k}{\partial \theta_j} P_k^{-1} + I_k \frac{\partial P_k}{\partial \theta_i}\left(\frac{\partial \hat{X}_{k|k-1}}{\partial \theta_j}\right)' H' P_k^{-1}$$

$$\left. + I_k^2 - \frac{1}{2} P_k \frac{\partial P_k}{\partial \theta_i}\frac{\partial P_k}{\partial \theta_j} P_k^{-2} + \frac{1}{2}(I_k^2 - P_k)\frac{\partial^2 P_k}{\partial \theta_i \partial \theta_j} P_k^{-1}\right\} P_k^{-1} \quad (4)$$

Calculation of the exact Hessian requires recursive computation of the second derivative matrices $\frac{\partial^2 V}{\partial \theta_i \partial \theta_j}$, $\frac{\partial^2 X_{k|k-1}}{\partial \theta_i \partial \theta_j}$, $\frac{\partial^2 P_k}{\partial \theta_i \partial \theta_j}$. Storage and computer time required are proportional to the square of the number of parameters, and in the present context these additional computation are not economically feasible.

Therefore the Hessian is calculated via the following approximation.

The approximation is based on the assumptions that the Hessian is nearly constant over the region of interest and that we may approximate the Hessian by its expected value over the underlying probability space. The two assumptions lead to the following positive definite expression for the Hessian involving first derivatives only.

$$\frac{\partial^2 L}{\partial\theta_i \partial\theta_j} = \sum_{i=1}^{N} \sum_{k=1}^{T} \left[H \left\{ E \frac{\partial \hat{X}_{k|k-1}}{\partial\theta_i} \quad \frac{\partial \hat{X}_{k|k-1}}{\partial\theta_j}' \right\} H' P_k^{-1} + \frac{1}{2} \frac{\partial P_k}{\partial\theta_i} \quad \frac{\partial P_k}{\partial\theta_j} \quad P_k^{-2} \right] \tag{5}$$

When the likelihood surface is approximately quadratic, then such an approximation works well, expecially near the maximum likelihood estimate. However, when this technique was tried in conjunction with the modified Newton-Raphson minimizing program MINIM [57] it failed. It was determined that the failure occured either because the problem was overfitted (hence unidentifiable) or because of the complex behavior of the likelihood surface, so that an estimate of the curvature at the desired point $\hat{\theta}$ did not adequately describe the properties of the surface. There may also be regions of the surface which do not have positive definite curvature, so that the positive definite Hessian estimate of the curvature is quantitatively and qualitatively false.

It is possible to improve the approximation by including all terms of (4) that involve no second derivatives, so that the calculation requires no new recursive calculations. In this approximation,

$$\frac{\partial^2 L}{\partial\theta_i \partial\theta_j} \sum_{i=1}^{N}\sum_{k=1}^{T}\left\{ H\left|\frac{\partial\hat{X}_{k|k-1}}{\partial\theta_i}\right|\left|\frac{\partial\hat{X}_{k|k-1}}{\partial\theta_j}\right|' H' P_k^{-1} + \frac{1}{2}\frac{\partial P_k}{\partial\theta_i}\frac{\partial P_k}{\partial\theta_j} P_k^{-2}\right.$$

$$+ (I_k^2 - P_k)\frac{\partial P_k}{\partial\theta_i}\frac{\partial P_k}{\partial\theta_j} P_k^{-3} + I_k H\left|\frac{\partial\hat{X}_{k|k-1}}{\partial\theta_i}\right|\frac{\partial P_k}{\partial\theta_j} P_k^{-2}$$

$$\left. + I_k \frac{\partial P_k}{\partial\theta_i}\left|\frac{\partial\hat{X}_{k|k-1}}{\partial\theta_j}\right|' H' P_k^{-2}\right\} \qquad (5)$$

The additional (third, fourth, fifth) terms in (5) have expected value zero when $\theta = \mathring{\theta}$. The term involving $(I_k^2 - P_k)$ may cause lack of positive definiteness when $P_k > I_k^2$. The last two terms vanish when the computed process I_k is white, but contribute otherwise. Only limited experimentation has been carried out with (5).

In practice the problem with bad behavior of the likelihood surface has been attacked by using the simple gradient method for the first few iterations in order to locate the ML estimates approximately. Then the Newton-Raphson method is used for final convergence. The method has been generally successful except when the Hessian matrix is near singular and second order methods fail. Robustness of convergence is a valuable (but subjective) criterioin in model identification. Relation to overfitting is discussed in Section 3.6.

APPENDIX 2

Calculation of Filter Gain Gradient

This appendix presents calculations for the gradient of the Kalman filter gain and the innovations variance, necessary as intermediates in order to obtain the gradient and Hessian matrix.

Differentiation of equation (3.2.2c) yields:

$$\frac{\partial K_{k+1}}{\partial \theta} = \frac{\partial V_{k+1|k}}{\partial \theta} H' P_{k+1}^{-1} - V_{k+1|k} H' P_{k+1}^{-2} \left[H \frac{\partial V_{k+1|k}}{\partial \theta} H' + \frac{\partial \sigma^2_v}{\partial \theta} \right]$$

Where $P_{k+1} = HV_{k+1|k}H' + \sigma^2_v$, (1)

and the index on θ_i has been suppressed.

To implement (1) we require a formula for $\frac{\partial V_{k+1|k}}{\partial \theta}$, obtained from equation (3.2.2d) by differentiating

$$\frac{\partial V_{k+1|k}}{\partial \theta} = \frac{\partial F}{\partial \theta} V_k F' + F V_k \frac{\partial F'}{\partial \theta} + \left[\frac{\partial G}{\partial \theta} G' + G \frac{\partial G'}{\partial \theta} \right] \sigma^2_w$$

$$+ F \frac{\partial V_k}{\partial \theta} F' + GG' \frac{\partial \sigma^2_w}{\partial \theta} \qquad (2)$$

Then (2.2.2e) yields

$$\frac{\partial V_k}{\partial \theta} = \left| \frac{\partial V_{k|k-1}}{\partial \theta} \right| - \left| \frac{\partial V_{k|k-1}}{\partial \theta} \right| H' H V_{k|k-1} + V_{k|k-1} H' H \left| \frac{\partial V_{k|k-1}}{\partial \theta} \right|' P_k^{-1}$$

$$+ V_{k|k-1} H' H V'_{k|k-1} P_k^{-2} \left[H \frac{\partial V_{k|k-1}}{\partial \theta} H' + \frac{\partial \sigma^2_v}{\partial \theta} \right] \qquad (3)$$

Together (1), (2), and (3) recursively define the filter gain gradient. These equations hold in general, for any definition of the parameter vector θ.

In the program the parameter vector is assumed to be

$$\theta' = [A_1, A_2, .., A_r, G_2, G_3, .., G_{r-1}, \sigma_w^2, \sigma_v^2, Q_{11}, Q_{12}, ..., Q_{1s}, Q_{22},$$
$$Q_{22}, Q_{23}, .., Q_{2s}, ..., Q_{ss}],$$

where r = model order,
and s = dimension of state vector.

When random trend terms are <u>not</u> present, r=s and $Q=V_{1|0}$.
When these terms <u>are</u> present then s=r+1 and

$$Q = \begin{bmatrix} V_{1|0} & D_R \\ D_R' & \sigma_R^2 \end{bmatrix}$$

This formulation is discussed in Section 2.5.

F and G are simple functions of ARMA parameter only and the following differentiation rules hold:

$$\frac{\partial F_{mn}}{\partial \theta_i} = \begin{cases} \delta(m,r)\ \delta(n,r+1-i) & i \leq r \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial G_m}{\partial \theta_i} = \begin{cases} \delta(m,i-r) & r < i \leq 2r-1 \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial \sigma_w^2}{\partial \theta_i} = \begin{cases} 1 & i=2r \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial \sigma_v^2}{\partial \theta_i} = \begin{cases} 1 & i=2r+1 \\ 0 & \text{else} \end{cases}$$

These rules make computation of the filter gradient relatively simple. It is done in conjunction with computation of the filter gain, as a precalculation performed once each iterative over all cases.

APPENDIX 3 An Analytic Example: Pre/post Treatment Design

In this appendix we apply the Kalman filter method analytically to the conventional pre/post treatment design. In this design, perhaps the commonest in educational evaluation, members of a treatment group and a control group are given a single "pretest" before the intervention and a single "posttest" afterwards. As demonstrated in Chapter III, two measurements are not sufficient for identifiability even of the simplest first order design. This is because the first order model involves four scalar parameters (A, σ^2_w, σ^2_v, $V_{1|0}$) while there are only three sufficient statistics in the covariance matrix among the measurements. Therefore this system is underidentified.

Nevertheless, the design is the most studied. Methodologists solve the problem of underidentification by making the additional assumption either that (1) measurement error is negligible, or (2) that it is known from other studies. Usually measurement error is expressed in terms of reliability, defined as the ratio of "true" to total variance. The reliability is incorrectly assumed to be a basic property of the instrument and is generally measured as Cronbach's α, a measure of consistency among test items.

Since by its definition reliability (generalizability) is a function of the population and the experimental design, which specify the "true" variance, this assumption is questionable. Another assumption that must be made is that $\sigma^2_R=0$, since estimation of terms involving R (fan spread) require two additional parameters.

While these assumptions are necessary if the identification is to be done at all, they introduce bias of unknown direction and magnitude.

The model may be written as follows:

$$\begin{aligned} X_{k+1} &= AX_k + W_k + U_k \\ Z_k &= X_k + V_k \end{aligned} \qquad (1)$$

Dynamic Equations

(1) First, the filter is set up. The Kalman filter variance and gain are calculated as follows. (Notice that H=1, G=1 and that all quantities are scalars.) The particular equation used from among the set (3.2.2) is indicated by symbols (i) to (v).

$$K_1 = \frac{V_{1|0}}{V_{1|0} + \sigma_v^2} \qquad \text{(iii)}$$

$$V_1 = \frac{\sigma^2_v}{V_{1|0} + \sigma_v^2} V_{1|0} \qquad \text{(v)}$$

$$V_{2|1} = \frac{A^2 \sigma_v^2}{V_{1|0} + \sigma_v^2} V_{1|0} + \sigma_w^2 \qquad \text{(iv)}$$

$$K_2 = \frac{V_{2|1}}{V_{2|1} + \sigma_v^2} \qquad \text{(iii)} \qquad (2)$$

Computation of Filter Gain

(2) Now the filter is applied to measurements in the following sequence of steps. It is assumed that the measurements are corrected for zero mean so that $E\{X_k\} = E\{Z_k\} = 0$.

Initial Condition: $\hat{X}_{1|0} = 0$

Step 1. Obtain Filtered State Estimate $\hat{X}_1$

$$\hat{X}_1 = \frac{V_{1|0}}{V_{1|0} + \sigma_v^2} Z_1 \qquad \text{(ii)}$$

Step 2. Obtain Forecast $\hat{X}_{2|1}$

$$\hat{X}_{2|1} = A X_1 = \frac{AV_{1|0}}{V_{1|0} + \sigma_v^2} Z_1 \qquad \text{(i)}$$

Step 3. Obtain Filtered State Estimate $\hat{X}_2$

$$\hat{X}_2 = \frac{\sigma^2 v}{V_{2|1} + \sigma_v^2} \frac{V_{1|0}}{V_{1|0} + \sigma_v^2} Z_1 + \frac{V_{2|1}}{V_{2|1} + \sigma_v^2} Z_2 \qquad \text{(ii)}$$

(3) Now we will use the output of the filter to set up the likelihood function. The innovations are given by

$$I_1 = Z_1$$

$$I_2 = Z_2 - \hat{X}_{2|1} = Z_2 - \frac{AV_{1|0}}{V_{1|0} + \sigma_v^2} Z_1 \qquad \text{(iii)}$$

The associated variances are:

$$P_1 = V_{1|0} + \sigma^2_v$$

$$P_2 = V_{2|1} + \sigma^2_v = \frac{A^2\sigma^2_v}{V_{1|0} + \sigma^2_v} V_{1|0} + \sigma^2_w + \sigma^2_v$$

Thus we can write the likelihood function using (3.2.5).

$$L = [Z_1^2/P_1 + \left[Z_2 - \frac{AV_{1|0}Z_1}{V_{1|0} + \sigma^2_v}\right]^2/P_2 + \log P_1 + \log P_2],$$

where the sum is over cases and P_1 and P_2 have the functional forms displayed above. Estimates for the unknowns $A, \sigma^2_w, \sigma^2_v, V_{1|0}$ are to be obtained by minimizing L over these parameters. Now these parameters occur in the equation for L in just <u>three</u> combinations: P_1, P_2 and $\frac{AV_{1|0}}{V_{1|0} + \sigma^2_v}$. Therefore $(A, \sigma^2_w, \sigma^2_v, V_{1|0})$ are not uniquely determined by minimizing L, i.e., they are under identified (as indicated by Theorem 1.)

In order to identify the system we must <u>specify</u> one relation involving the parameters. As we have discussed, the choice conventionally made in evaluation theory is <u>either</u> $\sigma^2_v = 0$ <u>or</u> $\frac{V_{1|0}}{V_{1|0} + \sigma^2_v} = \rho$, a known parameter describing the test. We will assume the latter, identifying as the reliability (generalizability) of the test. Then we obtain the following equations for $\hat{A}$ and $\hat{V}_{1|0}$ by differentiating L:

$$\hat{A} = \frac{1}{\rho} \frac{\sum z_2 z_1}{\sum z_1^2}$$

$$\hat{V}_{1|0} = \rho \frac{\sum z_1^2}{N}$$

The estimate $\hat{A}$ is the regression of Z_2 on Z_1, disattenuated by the reliability ρ. (This correction was suggested by Porter [50].) The estimate $\hat{V}_{1|0}$ is the estimated initial true score variance obtained as the reliability. Thus we can see that the identification technique automatically applies the disattenuation formula, which was to be expected since it is a modified MLE technique. In more complicated cases, beyond hand calculation, these measurement effects are also correctly treated.

Now let us solve the identification problem for the parameters $\bar{X}_{1|0}$ and U_1, by minimizing L_{II}. The filter is still defined by equations (2). Application of the filter to mean data $\bar{Z}_k$ yields:

$$\hat{\bar{X}}_1 = \frac{\sigma_v^2}{V_{1|0} + \sigma_v^2} \hat{\bar{X}}_{1|0} + \frac{V_{1|0}}{V_{1|0} + \sigma_v^2} \bar{Z}_1$$

$$\hat{\bar{X}}_{2|1} = A\hat{\bar{X}}_1 + U_1$$

The innovations are

$$I_1 = (\bar{Z}_1 - \hat{\bar{X}}_{1|0})$$

$$I_2 = (\bar{Z}_2 - A\hat{\bar{X}}_{2|1} - U_1)$$

$$= (\bar{Z}_2 - \frac{A\sigma_v^2}{V_{1|0} + \sigma_v^2} \hat{\bar{X}}_{1|0} - \frac{AV_{1|0}}{V_{1|0} + \sigma_v^2} \bar{Z}_1 - U_1)$$

The innovations may be made to vanish by setting

$$\hat{\bar{X}}_{1|0} = \bar{Z}_1, \text{ and}$$

$$U_1 = \bar{Z}_2 - A\bar{Z}_1,$$

thus minimizing L_{II} to 0.

Therefore our estimate for a treatment effect U_1 is obtained by the following steps:

(1) Regress Z_2 on Z_1 to obtain $\hat{A}$, uncorrected.

(2) Disattenuate A by division by generalizability ρ.

(3) Use the resulting estimate to form

$$\hat{U}_1 = \bar{Z}_2 - \hat{A}\bar{Z}_1.$$

APPENDIX 4 Proof of Main Identifiability Theorem

Preliminaries

In this appendix we prove that the structurally specified matrices $(F,G,P,Q,V_{1|0})$ of the discrete time nonstationary stochastic dynamic system

$$
\begin{aligned}
&X_{k+1} = FX_k + GW_k \\
&Z_k = HX_k + V_k \\
&(\forall k)\, E[X_1, W_k, V_k] = 0 \\
&(k,m)\, E\begin{bmatrix} X_1 \\ W_k \\ V_k \end{bmatrix} [X_1' W_m' V_m'] = \begin{bmatrix} V_{1|0} & 0 & 0 \\ 0 & P\delta_{km} & 0 \\ 0 & 0 & Q\delta_{km} \end{bmatrix}
\end{aligned}
\tag{1}
$$

are identifiable (T,Ω), under certain conditions to be stated. Recall that a parameter vector is second order identifiable (T,Ω) when the map $\tilde{\Phi}: \Omega \to \mathbb{R}^{Tq(Tq+1)/2}$ that carries the point θ to the covariance Σ_T of sequences $\{Z_k; k=1,2,..,T\}$ is one-to-one. The sequence length T must equal or exceed a certain minimum value to be specified, depending on conditions on $V_{1|0}$, the initial state covariance matrix. We will require that points in Ω satisfy stability and inverse stability conditions to be defined.

The structures of matrices in (1) are specified to obtain uniqueness. Required structures and dimensional specifications are as follows, where the notation $M_1 \otimes M_2$ indicates the Kronecker product of M_1 and M_2.

$$
\begin{aligned}
G &= \left[G_r ; G_{r-1} ; \ldots ; G_1 \right] \quad \text{where } G_1 = I_q \\
H &= \left[0, 0, \ldots, 0, 1 \right] \otimes I_q \\
A' &= \left[A_1, A_2, \ldots, A_r \right] \otimes I_q = \underline{A}' \otimes I_q \\
F &= \begin{bmatrix} A_1 & A_2 & \cdots & A_r \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 \; 1 & 0 \end{bmatrix} \otimes I_q = f \otimes I_q
\end{aligned}
\tag{2}
$$

Matrix Structures

X_k	qr x 1
$F, V_{1\|0}$	qr x qr
G, H', A	qr x q
W_k, Z_k, V_k	q x 1
G_i, Q, P	q x q
A_i	1 x 1

(3)

Dimensional Specifications

All matrices are divided into q x q blocks that form the basis of our indexing system, which originates in the lower right corner. Thus $[M]_{ij}$ is the (i,j)'th q x q block of M.

In light of the specifications (2), (3) we can define the parameter θ as a vector

$$\theta = (\underline{F},\underline{G},\underline{P},\underline{Q},\underline{V_{1|0}}) \tag{4}$$

The notation $\underline{M}$ indicates an ordered set of independent parameters specifying the matrix M. For instance $\underline{F}=(A_1,A_2,..,A_r)$, while $\underline{P}=(P_{11},P_{12},P_{22},..,P_{qq})$. We will write the total dimension of the vector as m. Then θ is a vector in the cone $\underline{R}^m$ obtained from R^m by restricting $P,Q,V_{1|0}$ to positive semi-definite matrices.

The matrix structure (2) is a minor variation of standard observable form (Brockett, [76], pp.105-109). Brockett shows that any matrix of rational functions such that the degree of the denominator of each element exceeds the degree of the numerator of that element can be represented in the form

$$H(Iz^{-1}-F)^{-1}G,$$

with F,G,H in the form specified by (2). The realization is <u>observable</u>, i.e., the observability matrix,

$$\mathcal{O}_k = \begin{bmatrix} HF^{k-1} \\ HF^{k-2} \\ \vdots \\ HF \\ H \end{bmatrix}, \tag{5}$$

has rank rq for $k \geq r$. In fact we have the following simple and useful algebraic form for $\mathcal{O}_{r+1}$:

$$\mathcal{O}_{r-1} = \begin{bmatrix} A' \\ I \end{bmatrix} = \begin{bmatrix} F \\ H \end{bmatrix} \tag{6}$$

One consequence of (6) is that the matrices $\{G_i; i=1,2,\ldots,r\}$ appear as the first r coefficient matrices in the Laurent expansion

$$H(Iz^{-1}-F)G = Iz + G_2z^2 + G_3z^3 + \ldots + G_rz^r + \ldots$$

It is convenient to denote the general coefficient in this series as G_i, so that

$$G_i = HF^{i-1}G, \qquad i=1,2,\ldots \infty .$$

where G' continues to represent the vector $(G_r, G_{r-1}, \ldots, G_1)$.

It must be stressed that the realization (2) is not in general controllable, and for q>1 is in fact nonminimal. The controllability matrix is given by the following.

$$\mathcal{C}(F,G) = [F^{r-1}G, \ldots, FG, G] = \begin{bmatrix} G_{2r-1} & \cdots & G_{r+1} & G_r \\ \vdots & & & \vdots \\ \vdots & & & G_3 \\ G_{r+1} & \cdots & G_3 & G_2 \\ G_r & \cdots & G_2 & I \end{bmatrix}, \tag{7}$$

The McMillan degree of (1) equals rank $\mathcal{C}(F,G) \leq rq$, so that when $\mathcal{C}(F,G) < rq$ our representation is not minimal. To specify a minimal realization for this problem would require specification of Kronecker indices and a different parameterization for each set of indices. In the single input case there is a single Kronecker index, equal to the McMillan degree, and this structural problem is obviated. In the q-input case we avoid the problem by paying the penalty of non-minimality.

Algebraic Form of $\Psi(\theta)$

The algebraic equation for the map

$$\boxed{\begin{aligned} &\Phi : \mathbb{R}^m \longrightarrow \mathbb{R}^{Tq(Tq+1)/2} \\ &\Sigma_T = \Phi(\theta) \end{aligned}} \tag{8}$$

is easily determined. Equations (1) give us the equations for the covariance Σ_T,

$$[\Sigma_T]_{ij} = EZ_i Z_j' = H(EX_i X_j')H' + \delta_{ij} Q, \tag{9}$$

where $$X_i = F^{i-1} X_1 + \sum_{k=1}^{i-1} F^{k-1} G W_{i-k}. \tag{10}$$

Assuming $i \geq j$, we can compute the following, using (10) and the last equation of (1).

$$EX_i X_j' = F^{i-1} V_{1|0} F'^{,j-1} + \sum_{k=1}^{j-1} F^{k-1+i-j} GPG'F'^{,k-1}$$

$$= F^{i-1}\left(V_{1|0} - \sum_{k=0}^{\infty} F^k GPG'F'^{,k}\right) F'^{,j-1} + \sum_{k=0}^{\infty} F^{k+i-j} GPG'F'^{,k} \tag{11}$$

In the latter form, we have added and subtracted a convergent infinite series that occurs in the stationary problem, assuming that F is stable. Therefore, substituting this result into (9), we obtain

$$[\Sigma_T]_{ij} = HF^{i-1}(V_{1|0} - \sum_{k=0}^{\infty} F^k GPG'F'^k)F'^{j-1}H$$

$$+ H \sum_{k=0}^{\infty} F^{k+i-j} GPG'G'^k H' + \delta_{ij} Q$$

This result can be conveniently reexpressed in the following form,

$$\Sigma_T = \mathcal{O}_T \tilde{V}_{1|0} \mathcal{O}_T' + \Lambda_T + I_T \otimes Q, \tag{12}$$

while $$\tilde{V}_{1|0} = V_{1|0} - \sum_{k=0}^{\infty} F^k GPG'F'^k, \tag{13}$$

$$\text{and } \Lambda_k = \begin{bmatrix} L_0 & L_1 & \cdots & L_2 & L_{k-1} \\ L_1' & L_0 & \ddots & \ddots & \vdots \\ L_2' & \ddots & \ddots & \ddots & L_2 \\ \vdots & \ddots & \ddots & \ddots & L_1 \\ L_{k-1}' & \cdots & L_2' & L_1' & L_0 \end{bmatrix}, \tag{14}$$

$$\text{where } L_i = H(\sum_{k=0}^{\infty} F^{k+i} GPG'F'^k)H'. \tag{15}$$

Equation (12) partitions the covariance matrix Σ_T into two parts: $\Lambda_T + I_T \otimes Q$ is a Toeplitz matrix equal to the steady state covariance, while $\mathcal{O}_T \tilde{V}_{1|0} \mathcal{O}'_T$ is the transient component. The matrix $\tilde{V}_{1|0}$ is the deviation of the initial state covariance $V_{1|0}$ from its steady state value $\sum_{k=0}^{\infty} F^k G P G' F'^k$.

In the next few pages we will construct an inverse function Φ^{-1} that maps $\Phi(\Omega)$ onto Ω, where Ω and the sequence length T are to be determined. The construction is carried out in a series of steps formalized as lemmas. When it is complete we will have determined (Ω,T) such that Φ is one-to-one and hence that $\mathring{\theta} \in \Omega$ is identifiable (T,Ω). These results will then be brought together in the statement of Theorem 1.

Determination of F, $\tilde{V}_{1|0}$ as Functions of Σ_T

Let S_k indicate the backward time shift matrix, i.e.,

$$S_k = \underset{kq \times kq}{\begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ I & 0 & & & \vdots \\ & I & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I & 0 \end{bmatrix}}. \tag{16}$$

and let the following semi-infinite matrices be defined.

$$\boxed{\begin{aligned} \Lambda &\triangleq \Lambda_\infty \\ S &\triangleq S_\infty \\ \Sigma &\triangleq \Sigma_\infty \\ \mathcal{O} &\triangleq \mathcal{O}_\infty \end{aligned}} \tag{17}$$

The action of the matrix S on a semi-infinite matrix M is as follows.

$$[SMS']_{ij} = [M]_{i+1,j+1}. \tag{18}$$

We introduce the notational convention

$$\Delta M \triangleq M - SMS' \tag{19}$$

so that Δ is a shift-difference operator. Then for any Toeplitz matrix M, such as Λ, $\Delta M = 0$.

Applying Δ to the equation

$$\Sigma = \Phi \tilde{V}_{1|0} \Phi' + \Lambda + I_{\infty} \otimes Q \tag{21}$$

(obtained from (12) by letting $T \longrightarrow \infty$), we get

$$\Delta\Sigma = \Phi \tilde{V}_{1|0} \Phi' - S\Phi \tilde{V}_{1|0} \Phi' S', \tag{22}$$

since Λ and $I_{\infty} \otimes Q$ are Toeplitz. It can easily be shown that

$$S\Phi = \Phi F. \tag{23}$$

Therefore we find that $\Delta\Sigma = \Phi(\tilde{V}_{1|0} - F\tilde{V}_{1|0}F')\Phi'$ (24)

$$= \Phi(V_{1|0} - FV_{1|0}F' - GPG')\Phi'$$

We can solve (24) for F and $\tilde{V}_{1|0}$ provided certain conditions are satisfied. Taking the lower right (r+1)x(r+1) blocks of (24) and using (6), we obtain.

$$\begin{bmatrix} A' \\ I \end{bmatrix} (\tilde{V}_{1|0} - F\tilde{V}_{1|0}F')[A \; I] = \Delta\Sigma_{r+1} \tag{25}$$

Thus $$\tilde{V}_{1|0} - F\tilde{V}_{1|0}F' = \Delta\Sigma_r \tag{26}$$

$$A'\Delta\Sigma_r = \underline{\Delta\Sigma'}_{r+1}, \tag{27}$$

where
$$\begin{bmatrix} \underline{\underline{\Delta\Sigma}}_{r+1} & \underline{\Delta\Sigma}'_{r+1} \\ \underline{\Delta\Sigma}_{r+1} & \Delta\Sigma_r \end{bmatrix} = \Delta\Sigma_{r+1}. \tag{28}$$

Equation (27) can be solved for A,

$$A = (\Delta\Sigma_r)^{-1} \underline{\Delta\Sigma}_{r+1}, \tag{29}$$

provided that $\Delta\Sigma_r$ is invertible. This fact motivates the following definition.

Definition The system described by the parameter $\theta = (\underline{F}, \underline{G}, \underline{P}, \underline{Q}, \underline{V}_{1|0})$ is said to be in full rank disequilibrium iff $|\Delta\Sigma_r| \neq 0$, (By (24), and the identity $\mathcal{O}_r = I$, we have

$$\Delta\Sigma_r = V_{1|0} - FV_{1|0}F' - GPG'$$

Then we immediately obtain

Lemma 1 If the system described by θ is in full rank disequilibrium, then $\underline{F}$ is a single valued function of Σ_{r+2}, namely the function defined by (29).

Furthermore:

Lemma 2 Let F be asymptotically stable. Then Equation (26) defines $\tilde{V}_{1|0}$ as a single valued function of F and Σ_{r+1}:

$$\tilde{V}_{1|0} = \sum_{k=0}^{\infty} F^k \Delta\Sigma_r F'^k. \tag{30}$$

If the system is in full rank disequilibrium, then $\tilde{V}_{1|0} > 0$.

Proof This is a result from Lyapunov theory. See Kalman [79]. Complete disequilibrium implies that $\Delta\Sigma_r > 0$.

When full rank disequilibrium does not hold, Lemma 1 fails, but a second condition for determination of F (and hence $\tilde{V}_{1|0}$) can be stated.

Lemma 3 Let F be nonsingular and $(F, V_{1|0}H')$ be controllable. Then F is a single valued function of Σ_{2r+1}.

Proof The proof follows the general lines of Ho and Kalman [64]. First observe that the following Hankel matrices can be constructed by arrangement of qxq blocks $HFV_{1|0}H'$, $HF^2,\ldots,HF^{2r}V_{1|0}H'$ from Σ_{2r+1}.

$$M_1 = \begin{bmatrix} HF^{2r-1}V_{1|0}H' & HF^{2r-2}V_{1|0}H' & \cdots & HF^{r}V_{1|0}H' \\ HF^{2r-2}V_{1|0}H' & HF^{2r-3}V_{1|0}H' & & \vdots \\ \vdots & & & HF^{2}V_{1|0}H' \\ HF^{r}V_{1|0}H' & \cdots & HF^{2}V_{1|0}H' & HFV_{1|0}H' \end{bmatrix} \quad (32)$$

$$M_2 = \begin{bmatrix} HF^{2r}V_{1|0}H' & HF^{2r-1}V_{1|0}H' & \cdots & HF^{r+1}V_{1|0}H' \\ HF^{2r-1}V_{1|0}H' & & & \vdots \\ \vdots & & & HF^{3}V_{1|0}H' \\ HF^{r+1}V_{1|0}H' & \cdots & HF^{3}V_{1|0}H' & HF^{2}V_{1|0}H' \end{bmatrix} \quad (33)$$

$$\text{Then } M_1 = \begin{bmatrix} HF^{r-1} \\ \vdots \\ HF \\ H \end{bmatrix} \left[F^r V_{1|0} H', \ldots, F V_{1|0} H' \right] \tag{34}$$

$$= \mathcal{O}_r F \mathcal{C}(F, V_{1|0} H')$$

where $\mathcal{O}_r$ is the observability matrix and $\mathcal{C}(F, V_{1|0}H')$ is the controllability matrix of F and $V_{1|0}H'$. We exploit the simple identity (6): $\mathcal{O}_r = I$,

$$M_1 = F\mathcal{C}(F, V_{1|0}H') \tag{35}$$

and by a similar calculation,

$$M_2 = F^2\mathcal{C}(F, V_{1|0}H') \tag{36}$$

By the conditions of the lemma, M_1 is nonsingular and therefore (uniquely)

$$F = M_2 M_1^{-1}. \tag{37}$$

Remark 1 We draw attention to the requirement that F be nonsingular. This is a necessary requirement in view of Lemma (2.3.1) which in state space representation, implies the following: When F is

singular, G,P,Q are <u>not</u> locally uniquely determined by the stationary covariance matrix $\Lambda_T + I_T \otimes Q$ and hence are not determined by the whole covariance matrix Σ_T.

<u>Remark 2</u> Suppose that the system is in equilibrium, so that

$$V_{1|0} = \sum_{k=0}^{\infty} F^k GPG'F'^k. \tag{38}$$

We will show that

$$\text{span } \mathcal{C}(F, V_{1|0}H') = \text{span } \mathcal{C}(F,G), \tag{39}$$

and hence $(F, V_{1|0}H')$ is controllable $\Leftrightarrow$ (F,G) is controllable $\Leftrightarrow$ (H,F,G) is minimal, the last implication holding since (H,F) is observable.

<u>Proof of Assertion (39)</u>

Let us denote the m'th block of the controllability matrix $\mathcal{C}(F, V_{1|0}H')$ by $\mathcal{C}_m(F, V_{1|0}H')$ so that by (38),

$$\mathcal{C}_m(F, V_{1|0}H') = \sum_{k=0}^{\infty} F^{k+m-1} GPG'F'^k H'. \tag{40}$$

Recall that F may be represented as the Kronecker product $f \otimes I_q$ where f is the rxr matrix

$$\begin{bmatrix} A_1 & A_2 & \cdots & A_r \\ 1 & 0 & & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{41}$$

Minimal polynomials for F and f are identical,

$$z^r - z^{r-1}A_1 - z^{r-2}A_2 - \ldots - A_r, \tag{42}$$

and imply that, evaluated at z=F,

$$\left[z^{k+1}, z^k, \ldots, z^{k-r}\right] = \left[z^k, z^{k-1}, \ldots, z^{k-r-1}\right] f', \tag{43}$$

which yields the relation

$$z^k = \left[z^{r-1}, z^{r-2}, \ldots, 1\right] f'^k \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \tag{44}$$

With suitable manipulation, and substitution of F for z, we obtain

$$F^{k+m-1} = \sum_{n=1}^{r} f'^{k+r-1}_{nm} F^{n-1}, \tag{46}$$

which (substituted into (40)) yields

$$\mathcal{C}_m(F, V_{1|0}H') = \sum_{n=1}^{r} F^{n-1}G \sum_{k=0}^{\infty} f'^{k+r-1}_{nm} PG'F'^k H' \tag{47}$$

Recognizing that $\mathcal{C}_n(F,G) = F^{n-1}G$, and rewriting (47) in Kronecker product form, we find

$$\mathcal{C}(F, V_{1|0}H') = \mathcal{C}(F,G) \sum_{k=0}^{\infty} f'^{k+r-1} \otimes (PG'F'^k H'), \tag{48}$$

which implies that

$$\text{Span } \mathcal{C}(F, V_{1|0}H') = \text{Span } \mathcal{C}(F,G) \tag{49}$$

if and only if the matrix

$$\sum_{k=0}^{\infty} f'^{k+r-1} \otimes (PG'F'^{k}H') \tag{50}$$

is invertible.

Since we have assumed that both f and P are invertible, (50) is invertible iff M^{-1} exists, where

$$M = \sum_{k=0}^{\infty} f^{k} \otimes HF^{k}G. \tag{51}$$

Since $f \otimes F$ is asymptotically stable, (51) can be rewritten in the form

$$M \triangleq (I_r \otimes H)(I - f \otimes F)^{-1}(I_r \otimes G). \tag{52}$$

Notice that since H and G are full rank, $I_r \otimes H$ and $I_r \otimes G$ are likewise full rank, inverse of $(I - G \otimes F)^{-1}$ exists. Therefore (52) defines a non-singular matrix and the assertion (39) is proved.

Determination of Remaining Parameters as Functions of Σ_T, F

We develop the necessary results as a sequence of lemmas.

Lemma 4 Let F be non-singular. Then the error covariance matrix Q is defined as a single valued function of F and Σ_{r+1}.

Proof First we construct the following qrxq matrix from the right hand q(r+1)xq block of Σ_{r+1}.

$$\begin{bmatrix} HF^{r}V_{1|0}H' \\ \vdots \\ HF^{2}V_{1|0}H' \\ HFV_{1|0}H' \end{bmatrix} = M \tag{53}$$

Then, using (5),

$$M = \Phi_r FV_{1|0}H' = FV_{1|0}H' \tag{54}$$

and since F is nonsingular, we can solve for $V_{1|0}H'$

$$V_{1|0}H' = F^{-1}M. \tag{55}$$

Now let us write the lower right qxq block of (12),

$$\Sigma_1 = H\tilde{V}_{1|0}H' + H\sum_{k=0}^{\infty} F^{k}GPG'F'^{k}H' + Q$$

$$= HV_{1|0}H' + Q$$

$$= HF^{-1}M + Q \text{ (using (55))},$$

so that $$Q = \Sigma_1 - HF^{-1}M \tag{56}$$

which exhibits Q as a function of F and Σ_{r+1}. Q.e.d.

Lemma 5 Let F be stable and non-singular. Then the spectral density matrix,

$$\Lambda(z) = \sum_{k=-\infty}^{\infty} L_k z^k, \tag{57}$$

is a function of Σ_{r+1}, F, and (of course) z.

Proof The spectral density matrix can be written as

$$\Lambda(z) = H(Iz^{-1}-F)^{-1}\bar{G} + (H(Iz-F)^{-1}\bar{G})' + H\bar{G}, \tag{58}$$

where $$G \triangleq \sum_{k=0}^{\infty} F^k GPG'F'^k H'. \tag{59}$$

Right multiplying (13) by H', a fixed matrix, and rearranging, we obtain

$$\sum_{k=0}^{\infty} F^k GPG'F'^k H' = V_{1|0}H' - \tilde{V}_{1|0}H'. \tag{60}$$

Now we use (59), (55), and (30) in (60) to obtain

$$\bar{G} = F^{-1}M - \sum_{k=0}^{\infty} F^k \Delta\Sigma_r F'^k H' \tag{61}$$

Substitution of (61) into (58) then results in an explicit formula for $\Lambda(z)$ as a function of F and Σ_{r+1}. Q.e.d.

The spectral density function (58) is that obtained by allowing the system

$$X_{k+1} = FX_k + GW_k$$

$$Z_k = HF_k$$

to achieve steady state stationarity. Thus determination of $\Lambda(z)$ in some sense converts the original problem to its stationary equivalent. Determination of (G,P) from $\Lambda(z)$ is a standard problem with a vast literature, from which we draw in the remainder of the development.

By simple algebraic manipulation it can be shown that the function

$$\Psi(z) = \sum_{k=0}^{\infty} HF^k Gz^k \tag{62}$$

is a factor of $\Lambda(z)$:

$$\Lambda(z) = \Psi(z) P \Psi'(z^{-1}) \tag{63}$$

Thus if $\Psi(z)$ and P can be determined uniquely from $\Lambda(z)$ we can solve Equation (61) for G, obtaining the last unknown parameters uniquely. The following theorem, adapted from Youla [75], as stated by Goodwin and Payne [74] specifies the conditions under which this is in fact the case.

Spectral Factorization Theorem Let $\Lambda(z)$ be a q x q discrete rational spectral density matrix with full normal rank. (A discrete rational spectral density matrix satisfies the following.

$$\Lambda(z^{-1}) = \Lambda'(z) \tag{64}$$

$$\Lambda(z) \geq 0 \text{ for } |z|=1$$

Since we require full normal rank,

$$\Lambda(z)>0, \ |z|=1, \text{ a.e.})$$

Then there exist a unique q x q rational matrix (z) and a unique q x q positive definite symmetric matrix P such that $\Lambda(z) = \Psi(z)P\Psi'(z^{-1})$ and

(1) $\Psi(0) = I$

(2) $\Psi(z)$ is analytic on the unit disk.

(3) The inverse matrix $\Psi^{-1}(z)$ is analytic inside the unit circle.

Now we can prove the main theorem, drawing together the results proved step by step up to this point.

Theorem 1 (Identifiability) Consider the map $\Phi : \Omega \to \mathbb{R}^{Tq(Tq+1)/2}$ generated by the dynamic model specified by (1)-(3). Suppose that each $\theta \in \Omega \subset \underline{\mathbb{R}}^M$ satisfies the following conditions.

(1) $\Psi^{-1}(z)$ is analytic for $|z|<1$, i.e., Ψ is inverse stable. By abuse of language we will also call θ inverse stable.

(2) F is asymptotically stable and nonsingular.

Let $\mathring{\theta}$ be an arbitrary point of Ω.

I. If the system represented by $\mathring{\theta}$ is in full rank disequilibrium, then $\mathring{\theta}$ is identifiable $(r+2, \Omega)$

II. If $(F, V_{1|0}H')$ is controllable, then $\mathring{\theta}$ is identifiable $(2r+1, \Omega)$.

Proof Assume that $\theta \in \Omega$, where Ω satisfies conditions (1) and (2).

(I) Assume that the system represented by θ is in complete disequilibrium, so that Lemma 1 applies, and F is a function of Σ_{r+2}. By lemmas 4 and 5, then Q and $\Lambda(z)$ are also functions of Σ_{r+2}. By the spectral density theorem, under the conditions stated, $\Psi(z)$ and P are uniquely determined by $\Lambda(z)$, and since

$$\Psi(z) = I + G_2 z + G_3 z^2 + .. + G_r z^{r-1} + \sum_{k=0}^{\infty} HF^k G z^k, \quad (49)$$

the vector G is a function of Σ_{2r+1}.

Finally,

$$V_{1|0} = \tilde{V}_{1|0} + \sum_{k=0}^{\infty} F^k GPG'F'^k, \quad (50)$$

and the entire vector $\theta = (\underline{F}, \underline{G}, \underline{P}, \underline{Q}, \underline{V}_{1|0})$ is determined as a single valued function of Σ_{r+2}. Thus, with T=2r+1, $\Phi : \Omega \to \mathbb{R}^{Tq(Tq+1)/2}$ is one-to-one and θ is identifiable $(r+2, \Omega)$.

(II) Assume that $(F, V_{1|0} H')$ is controllable. Then Lemma 3 applies and F can be uniquely determined from Σ_{2r+1}. The remainder of the proof is the same as for (I). θ is identifiable $(2r+1, \Omega)$. Q.e.d.

APPENDIX 5 Identifiability of Random Trend Parameters

This appendix extends the the proof of identifiability of the parameter vector $(\underline{F},\underline{G},\sigma^2_w,\sigma^2_v,\underline{V}_{1|0})$ to include the random trend parameters σ^2_R, D_R of the following model, where the case index i is made explicit.

$$X_{k+1,i} = FX_{ki} + GW_{ki} + CR_i$$

$$Z_{ki} = HX_{ki} + V_{ki}$$

$$(\forall k,i)\ E(X_{ki},W_{ki},V_{ki}) = 0$$

$$(\forall k,m,i,j)\ E\begin{bmatrix} X_{1i} \\ R_i \\ W_{ki} \\ V_{ki} \end{bmatrix} (X'_{1,j},R_j,W_{mj},V_{mj}) = \begin{bmatrix} V_{1|0} & D'_R & 0 & 0 \\ D_R & \sigma^2_R & 0 & 0 \\ 0 & 0 & P\delta_{km} & 0 \\ 0 & 0 & 0 & Q\delta_{km} \end{bmatrix} \delta_{ij} \qquad (1)$$

Matrices F,G,H have the structures specified in Appendix 4. In addition, we define

$$C' = (1,0,0,..,0).$$

In this appendix we deal only with the scalar case, corresponding to q=1 in Appendix 4. Thus dimensions are specified as follows (where the index i is again suppressed).

X_k, D_R, G, H', A	r x 1
$F, V_{1\|0}$	r x r
$W_k, Z_k, V_k, \sigma^2_w, \sigma^2_v, \sigma^2_R$	1 x 1

(2)

The parameter vector is $\theta_I = (\underline{F}, \underline{G}, \sigma^2_w, \sigma^2_v, \underline{V}_{1|0}, D_R, \sigma^2_R)$, where the underlines indicate "vectorization" of independent parameters. Thus $\underline{F}=A'$, $\underline{G}=(G_2, G_3, \ldots, G_r)$, and $\underline{V}_{1|0}$ consists of the upper triangle of $V_{1|0}$, arranged as a vector.

We have termed the model (1) the "random trend" model because each individual realization is forced by an input R that is constant for that realization (but random across realizations). Before we proceed with the theorems to be proved we will show equivalence of the random trend model to another we term the random origin model:

$$X_{k+1,i} = FX_{ki} + GW_{ki}$$

$$Z_{ki} = HX_{ki} + Z_{0i} + V_k \qquad (3)$$

Here each measurement is subject to a bias Z_{0i} that is constant for each individual i but random across individuals:

$$EZ_{0i} = 0$$

$$E\begin{bmatrix} X_{1i} \\ Z_{0i} \end{bmatrix} (X'_{1\ j}, Z_{0i}, W_{kj}, V_{kj}) = \begin{bmatrix} V_{1|0} & D'_0 & 0 & 0 \\ D_0 & \sigma^2_0 & 0 & 0 \end{bmatrix} \qquad (4)$$

Remaining statistics are the same as before. This model is parameterized by $(\underline{F}, \underline{G}, \sigma^2_w, \sigma^2_v, \underline{V}_{1|0}, D_0, \sigma^2_0)$.

Measurements for the random trend model have the following form (where, again, the index i is suppressed), obtained from the use of the variation of constants formula.

$$Z_k = H \sum_{i=1}^{k-1} F^{i-1} GW_{k-i} + HF^{k-1}(X_1 - (I-F)^{-1}GCR)$$

$$+ H(I-F)^{-1}GCR + V_k, \tag{5}$$

provided F is asymptotically stable.

If we set $X_1 - (I-F)^{-1}GCR \longrightarrow X'_1$

$$H(I-F)^{-1}GCR \longrightarrow Z_0, \tag{6}$$

and then suppress the prime on the new variable X'_1, equation (1) becomes:

$$Z_k = H \sum_{i=1}^{k-1} F^{i-1} GW_{k-i} + HF^{k-1}X_1 + Z_0 + V_k \tag{7}$$

which is a representation of Z_k in random origin form. We can move freely from the random trend parameterization to the random origin, and back again. We will use the random origin model in this section because of greater notational simplicity. The substantive content of the equivalence is that a model in which all measurements are relative to a randomly distributed subject-by-subject origin is equivalent to one with common measurement origin but with random subject-to-subject driving forces.

The algebraic form of the map $\Phi: \theta_I \longrightarrow \Sigma_T$, calculated from (7) by the technique used in deriving (A4.12), is

$$\boxed{\Sigma_T = \mathcal{O}_T \tilde{V}_{1|0} \mathcal{O}_T' + \sigma_v^2 I + \mathcal{A}_T + \mathcal{O}_T D_0 \mathbf{1}' + \mathbf{1} D_0' \mathcal{O}_T' + \sigma_0^2 \mathbf{1}\mathbf{1}'} \quad (8)$$

where $$\tilde{V}_{1|0} = V_{1|0} - \sum_{k=0}^{\infty} F^k G \sigma_w^2 G' F'^k, \quad (9)$$

$$\text{and } \mathcal{A}_T = \begin{bmatrix} L_o & L_1 & L_2 & & L_{T-1} \\ L_1 & L_o & \cdot & \cdot & \cdot \\ L_2 & \cdot & \cdot & \cdot & L_2 \\ & \cdot & \cdot & \cdot & L_1 \\ L_{T-1} & L_2 & L_1 & \cdot & L_o \end{bmatrix} \quad (10)$$

$$L_i = H\left(\sum_{k=0}^{\infty} F^{k+i} G G' F'^k\right) H' \quad (11)$$

$$\mathbf{1}' = (1, 1, \ldots, 1) \quad (12)$$

Note that equation (9) extends equation (A5.13) in a simple way: new terms involving D_0 and σ^2_0 appear only additively, so that we need only prove identifiability of these new variables to reduce the problem to the form discussed in Appendix 4.

We will prove the following lemma in an informal style, since a similar proof was given in Appendix 4 (Lemmas (A4.3), (A4.4)).

Lemma 1 Let $(F, V_{1|0} H')$ be controllable, F be invertible, and (I-F) be invertible. Then the quantities $V_{1|0} H'$, σ^2_v, $D'_0 H' + \sigma^2_0$, and F are functions of Σ_{2r+2}.

Proof The RH column of (9), in semi-infinite matrix form, is given by

$$\mathcal{O}V_{1|0}H' + \mathcal{I}(D_0'H'+\sigma_0^2) + \sigma_v^2\delta = \underline{\Sigma}, \tag{13}$$

$$\text{with } V_{1|0} = \tilde{V}_{1|0} + \sum_{k=0}^{\infty} F^k G \sigma_w^2 G' F'^k \tag{14}$$

and $\delta' = (\ldots,0,0,..,0,1)$.

Introducing the shift operator S defined by Equation (A4.16) we can show that $S\mathcal{O}=\mathcal{O}F$, while $S\mathcal{I}=\mathcal{I}$, and applying the matrix (I-S) from the left in Equation (13), we obtain

$$\mathcal{O}(I-F)V_{1|0}H' + \sigma_v^2\delta = (I-S)\underline{\Sigma} \tag{15}$$

Rows 2 to 2r+1 of Equation (15) assign values to the sequence $\{F^k(I-F)FV_{1|0}H';k=0,1,\ldots,2r+1\}$. The Ho-Kalman procedure used in the proof of Lemma (A4.3) can be employed to find F, provided that

$$(F,(I-F)FV_{1|0}H')$$

is <u>controllable</u>. But controllability of $(F,(I-F)FV_{1|0}H')$ follows from controllability of $(F,V_{1|0}H')$. Thus provided $(F,V_{1|0}H')$ is controllable, F is a function of Σ_{2r+2}.

The procedure also results in determination of $(I-F)FV_{1|0}H'$ as a function of Σ_{2r+2}. Thus we can compute $V_{1|0}H'$ uniquely <u>provided</u> that both F and (I-F) are invertible, i.e., that F does not have characteristic roots at the origin or <u>on</u> the unit circle.

With $V_{1|0}H'$ and F determined, one can determine $(D'_0H' + \sigma^2_0)$ and σ^2_v as a function of Σ_2, using (13). <u>Q.e.d.</u>

Now we will apply the same shift-difference technique used in Appendix 4 to obtain Equation (A4.24). If we define the operation Δ by

$$\Delta M \triangleq M - SMS',$$

then, since

$$S\mathcal{O} = \mathcal{O}F$$

$$S\mathbf{1} = \mathbf{1}$$

$$SIS' = I, \text{ where all matrices are semi-infinite,}$$

we have

$$\Delta\Lambda = 0$$

$$\Delta \sigma_0^2 \mathbf{1}\mathbf{1}' = 0.$$

and from (8), with $T = \infty$, we get

$$\Delta\Sigma = \mathcal{O}(\tilde{V}_{1|0} - F\tilde{V}_{1|0}F')\mathcal{O}' + \mathcal{O}(I-F)D_0\mathbf{1}' + \mathbf{1}D_0'(I-F')\mathcal{O}' \qquad (16)$$

Now, concentrating on the lower right $(r+1)\times(r+1)$ block of (16), and using

$$\mathcal{O}_{r+1} = \begin{bmatrix} A' \\ I \end{bmatrix}.$$

we obtain

$$(\Delta\Sigma)_{r+1} = \begin{bmatrix} A' \\ I \end{bmatrix} (\tilde{V}_{1|0} - F\tilde{V}_{1|0}F') \begin{bmatrix} A' & I \end{bmatrix} \qquad (17)$$

$$+ \begin{bmatrix} A' \\ I \end{bmatrix} (I-F)D_0\mathbf{1}' + \mathbf{1}D_0'(I-F') \begin{bmatrix} A & I \end{bmatrix}.$$

The lower right r x r block of (17) is:

$$(\Delta Z)_r = \tilde{V}_{1|0} - F\tilde{V}_{1|0}F' + (I-F)D_0 1' + 1D_0'(I-F') \quad (18)$$

Substituting (18) into the (r+1)'st row, right r columns of (16) yields, after manipulation,

$$D_0' = \frac{1}{(1 - \sum_{i=1}^{r} A_i)} (\underline{\Delta\Sigma}_{r+1} - A'\Delta\Sigma_r)(I-F')^{-1} , \quad (19)$$

where Σ_{r+1} is the (r+1)'st row, first r columns of $\Delta\Sigma_{r+1}$. Thus we obtain the following lemma.

Lemma 2 D_o and σ_o^2 are determined as functions of Σ_{2r+2} provided F is invertible and asymptotically stable, and that $(F, V_{1|0}H')$ is controllable.

Proof The proof has already been presented. Note that stability of F implies that $\sum_{k=1}^{r} A_k \neq 1$, $\exists (I-F)^{-1}$. Q.e.d.

Since D_0 and $\sigma^2{}_0$ are known functions of Σ_{2r+2}, we can convert Equation (8) to the form

$$\mathcal{O}_T V_{1|0} \mathcal{O}_T' + \Lambda_T + I\sigma_v^2 = \tilde{\tilde{\Sigma}}_T \quad (20)$$

where $\tilde{\tilde{\Sigma}}_T$ is determined as a function of Σ_T and Σ_{2r+2}. This is the same form as the problem of Appendix 4. Therefore, without further discussion we can assert the following.

Theorem 1 The vector $\theta_I = (\underline{F}, \underline{G}, \sigma^2{}_w, \sigma^2{}_v, V_{1|0}, D_0, \sigma^2{}_0)$ or equivalently, $(\underline{F}, \underline{G}, \sigma^2{}_w, \sigma^2{}_v, V_{1|0}, D^2{}_R, \sigma^2{}_R)$ is identifiable $(2r+2, \Omega)$ provided the following conditions hold throughout Ω.

(1) F is stable.
(2) θ is inverse stable.

For the particular point θ_I,
(1) $(F, V_{1|0}H')$ is controllable.
(2) F is invertible.

APPENDIX 6 Estimation of Mean Growth Curve

In this appendix we present a closed form solution to the problem of obtaining the estimate $\hat{\theta}'_{II} = (\hat{\bar{X}}_{1|0}, \{\hat{U}_k\})$ discussed in Section 3.4. It is assumed that the estimates $\hat{\theta}'_I = (A, G, \sigma^2_v, \sigma^2_w, V_{1|0})$ are known from the procedure described in Chapter III, and that the T-vector θ_{II} is constrained by the linear equations

$$A\theta_{II} = 0, \tag{1}$$

where A is a C x T matrix of full rank $C \leq T$. The solution algorithm forms a set of subroutines of the computer program (Appendix 8).

The state prediction updating formulas (3.4.4) can be combined into the single dynamic equation

$$\hat{\bar{X}}_{k+1|k} = F(I-K_kH)\,\hat{\bar{X}}_{k|k-1} + FK_k\,\bar{Z}_k + CU_k \tag{2}$$

which we assume to have transition matrix $\bar{\Phi}(k+p,k)$. The variation of constants formula can then be used to obtain $\hat{\bar{X}}_{k+1|k}$ as a linear function of $\theta'_{II} = (\bar{X}'_{1|0}, \bar{U}')$.

$$\bar{X}_{2|1} = \bar{\Phi}_{21}\,\bar{X}_{1|0} + CU_1 + FK_1\bar{Z}_1$$

$$\bar{X}_{3|2} = \bar{\Phi}_{31}\,\bar{X}_{1|0} + \bar{\Phi}_{32}\,(CU_1+FK_1\bar{Z}_1)+(CU_2+FK_2\,\bar{Z}_2)$$

...

$$\bar{X}_{T|T-1} = \bar{\Phi}_{T1}\,\bar{X}_{1|0} + \bar{\Phi}_{T2}(CU_1+FK\,\bar{Z}_1)+..+(CU_{T-1}+FK_{T-1}\,\bar{Z}_{T-1}) \tag{3}$$

Recall that $\{\bar{z}_k\}$ is the sequence of cross-sectional means. As discussed in Section 3.4, the estimate $\hat{\theta}_{II}$ is obtained by minimizing the sum of squares

$$L_{II} = N \sum_{k=1}^{T} (\bar{z}_k - H\hat{\bar{x}}_{k|k-1})^2 P_k^{-1} \tag{4}$$

over $\theta_{II} \in \Omega_{II}$, the domain defined by the constraints (1). Substituting the expressions (3) into (4) yields, after routine computation, the formula

$$L_{II}/N = \sum_{k=1}^{T} \left(S_k \bar{z} - E_k \begin{bmatrix} \tilde{\bar{x}}_{1|0} \\ U \end{bmatrix}\right)^2 P_k^{-1}, \tag{5}$$

where $Z' = (z_1, z_2, \ldots, z_T)$, and E_k is the k'th row of the matrix

$$E \triangleq \begin{bmatrix} H & & & \\ H\Phi_{21} & HC & 0 & \\ H\Phi_{31} & H\Phi_{32}C & HC & \\ \vdots & \vdots & & \\ H\Phi_{T1} & H\Phi_{T2}C & \cdots & HC \end{bmatrix}, \tag{6}$$

$T \times r$ $\quad$ $T \times (T-r)$

while S_k is the k'th row of the matrix

$$S \triangleq \begin{bmatrix} 1 & & & & \\ -HFK_1 & 1 & & & \\ -H\phi_{32}FK_1 & -HFK_2 & 1 & & \\ -H\phi_{42}FK_1 & -H\phi_{43}FK_2 & \cdot & \cdot & \\ \vdots & & \cdot & \cdot & \\ -H\phi_{T2}FK_1 & \cdot\ \cdot\ \cdot\ \cdot & -HFK_{T-1} & 1 & \end{bmatrix}. \tag{8}$$

The constraints $A\theta_{II}=0$ are adjoined to (5) to form the Lagrangian

$$\mathcal{L} = \bar{Z}'S'P^{-1}S\ \bar{Z} - 2\ Z'S'P^{-1}E\theta_{II} + \theta_{II}'\ E'P^{-1}E\theta_{II} + 2\lambda'A\theta_{II} \tag{9}$$

where $P = \text{diag}\ [P_1, P_2, \ldots, P_T]$.

Differentation with respect to θ_{II} yields:

$$E'\ P^{-1}E\theta_{II} + A'\lambda - E'P^{-1}S\ \bar{Z} = 0 \tag{10}$$

Assuming E^{-1} exists we may left multiply by $A(E'P^{-1}E)^{-1}$ and use (1) to obtain:

$$A(E'P^{-1}E)^{-1}A'\lambda = A(E'P^{-1}E)^{-1}E'P^{-1}\ S\ \bar{Z}$$

so that $$\lambda = (A(E'P^{-1}E)^{-1}A')^{-1} A(E'P^{-1}E)^{-1} E'P^{-1} S \bar{Z}. \quad (11)$$

Substituting (11) into (10) and manipulating, we obtain

$$\theta_{II} = (I - (E'P^{-1}E)^{-1} A'(A(E'P^{-1}E)^{-1}A')^{-1}A)E^{-1}S \bar{Z} \quad (12)$$

The expression (12) may then be substituted into (5) to obtain the minimized sum of squares

$$\underset{\theta_{II} \in \Omega_{II}}{\text{Min}} L_{II}(\theta_I, \theta_{II}) = \bar{Z}'S'E'^{-1}A'(A(E'P^{-1}E)^{-1}A')^{-1}AE^{-1}S\bar{Z} \quad (13)$$

Equation (13) is an explicit evaluation of the function (3.4.7) discussed in Section 3.4 and demonstrates that global maximization of $L(\theta_I, \theta_{II})$ is feasible. One would merely adjoin (13) to $L_I(\theta_I)$ and maximize the sum, as suggested by (3.4.8).

When there are no constraints, so that $\Omega_{II} = R^T$, (12) and (13) reduce to

$$\theta_{II} = E^{-1}S\bar{Z} \quad (14)$$

and $$\underset{\theta_{II} \in R^T}{\text{Min}} L_{II}(\theta_I, \theta_{II}) = 0,$$

a result used in the proof of Theorem 3.4.2. More generally, the rank of the matrix

$$S'E'^{-1}A'(A(E'P^{-1}E)^{-1}A')^{-1}AE^{-1}S$$

that occurs in (13) is equal to rank A=C. In this case the sum of squares is chi-square with C degrees of freedom.

In this formulation the constraint matrix A was considered known. In practice the form of A specifies structural assumptions about the sequence $\{U_k\}$, and we may wish to test a sequence of hypotheses in order to achieve structure determinations. Since (13) provides a chi-square goodness-of-fit test, such hypothesis tests are easily constructed. Hypotheses about treatment effects in educational evaluations also take this form.

APPENDIX 7 Local Identifiability and the Hessian Matrix

Preliminaries

This appendix provides proofs for the two theorems stated in Section 3.6.

Consider a sample of n sequences $\{z_k; k=1,2,\ldots,T\}$, normally distributed with (vector) mean 0 and covariance Σ_T. We assume that Σ_T is a function of an m x 1 parameter vector under the map

$$\underset{\sim}{\phi} : \Omega \longrightarrow \mathbb{R}^{T(T+1)/2}$$
$$\theta \longmapsto \Sigma, \tag{1}$$

where $m \leq T(T+1)/2$ and Σ is represented as a vector

$$\Sigma' \triangleq (\Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \ldots, \Sigma_{TT}), \tag{2}$$

formed from the upper triangle of Σ_T. We denote by $J(\theta) \triangleq \frac{\partial \phi}{\partial \theta}$ the $T(T+1)/2$ x m dimensional Jacobian matrix at point θ.

We say that $\underset{\sim}{\phi}$ has rank r at θ if rank $J(\theta)=r$, and will call θ a regular point if rank $J(\theta)$ is constant in some sufficiently small neighborhood of θ. We denote by $\mathcal{X}$ the set of all regular points and by $\mathcal{X}^c$ its complement.

Relation of Hessian and Jacobian Matrices

The LLF for a general normal process can be written

$$L = \frac{n}{2} [\log|\Sigma| + \operatorname{trace} \Sigma^{-1} S_n], \tag{3}$$

where S_n is the sample covariance matrix of n observed sequences $\{z_k; k=1,2,\ldots,T\}$. Let $[L_{ij}]_\theta$ represent the corresponding Hessian matrix, evaluated at θ. Then we have:

Lemma 1 The matrix $[L_{ij}]_\theta$ converges almost surely to the matrix M, where

$$M_{ij} = \frac{n}{2} \text{ trace } (\Sigma^{-1/2} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1/2})(\Sigma^{-1/2} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1/2}), \qquad (4)$$

and M is singular if and only if rank $\Phi < m$.

Proof By differentation of (4) we obtain

$$\frac{\partial L}{\partial \theta_i} = \frac{n}{2} \text{ trace } (\Sigma^{-1} S_n \Sigma^{-1} - \Sigma^{-1}) \frac{\partial \Sigma_T}{\partial \theta_i} \qquad (5)$$

$$L_{ij} = \frac{-n}{2} \text{ trace } [\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} S_n \Sigma^{-1}$$

$$+ \Sigma^{-1} S_n \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} - \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1}] \frac{\partial \Sigma}{\partial \theta_i} \qquad (6)$$

$$+ \frac{n}{2} \text{ trace } (\Sigma^{-1} S_n \Sigma^{-1} - \Sigma^{-1}) \frac{\partial^2 \Sigma}{\partial \theta_i \partial \theta_j}$$

(where we have suppressed the index of Σ_T). Since L_{ij} is a continuous function of S_n, and $\{S_n\}$ is a strongly consistent sequence of estimates for Σ, L_{ij} converges a.s. to the matrix obtained from (6) by substituting Σ for S_n, namely to M as cited in the statement of the theorem.

Clearly M is singular if and only if the Jacobian matrix $J(\mathring{\theta}) = \left.\frac{\partial \Sigma}{\partial \theta}\right|_{\mathring{\theta}}$ is singular, i.e., if rank $\underline{\Phi} < m$ at $\mathring{\theta}$. Q.e.d.

Sufficient Conditions for Local Identifiability

Theorem 1A Let $\underline{\Phi}$ have full rank m at $\mathring{\theta}$. Then there exist a neighborhood Ω of $\mathring{\theta}$ and a neighborhood Ω_z of $\underline{\Phi}(\mathring{\theta})$ such that $\underline{\Phi}$ maps Ω homeomorphically onto Ω_z (that is, $\underline{\Phi}$ is one-to-one onto Ω_z and both $\underline{\Phi}$ and $\underline{\Phi}^{-1}$ are continuous).

Proof We use a standard technique in order to apply the inverse function theorem. Assume that $\underline{\Phi}$ has full rank m at and that (without loss of generality) the first m rows of $J(\mathring{\theta})$ form an m x m nonsingular matrix. We define the map.

$$\underline{\Phi}' : \mathbb{R}^{T(T+1)/2} \longrightarrow \mathbb{R}^{T(T+1)/2}$$

$$\Phi_i'(\theta,\nu) = \Phi_i(\theta); \qquad i=1,2,\ldots,m$$

$$\Phi_i'(\theta,\nu) = \nu_i + \Phi_i(\theta); \qquad i=m+1,m+2,\ldots,T(T+1)/2$$

Since $\underline{\Phi}'$ is continuously differentiable and has rank $T(T+1)/2$, we can apply the inverse function theorem (Wallace [69], Theorem 2-1) to show that there are neighborhoods Ω' of $(\mathring{\theta},0)$ and Ω_z' of $\underline{\Phi}(\mathring{\theta})$ such that $\underline{\Phi}'$ maps Ω' homeomorphically onto Ω_z'. By construction, the map $\underline{\Phi} : \underline{\mathbb{R}}^m \longrightarrow \mathbb{R}^{T(T+1)/2}$ is the restriction of $\underline{\Phi}'$ to $\underline{\mathbb{R}}^m$, and $\underline{\Phi}(\underline{\mathbb{R}}^m)$ is a submanifold of $\mathbb{R}^{T(T+1)/2}$. Thus the statement of the theorem holds with $\Omega = \Omega' \cap \mathcal{M}$ and $\Omega_z = \Omega' \cap \underline{\Phi}'(\mathcal{M})$, where $\mathcal{M}$ is the set $\{\theta,\nu \mid \nu = 0\}$. Q.e.d.

By lemma 1, $[L_{ij}]$ converges to a nonsingular matrix M if and only if $\underline{\Phi}$ has rank m. Hence if $[L_{ij}] \longrightarrow M<0$, $\underline{\Phi}$ is locally one-to-one and $\mathring{\theta}$ is locally identifiable (T).

Necessary Conditions for Local Identifiability

Theorem 1B Let $\theta \in \mathcal{R}$ and let rank $\bar{\Phi} = P < m$ for all $\theta \in \Omega_1$, where Ω_1 is an open neighborhood of $\mathring{\theta}$. Then $\mathring{\theta}$ is not locally identifiable (T).

Proof (Suggested by Approach of Fisher [60], pp.163-167.) Assume that rank $\bar{\Phi} = P < m$ for all $\theta \in \Omega_1$. Without loss of generality we assume that the system of continuously differentiable equations

$$\bar{\Phi}_i(\theta) = \bar{\Sigma}_i; \qquad i=1,2,\ldots,P \tag{8}$$

has non-zero Jacobian determinant

$$\frac{\partial(\bar{\Phi}_1, \bar{\Phi}_2, \ldots, \bar{\Phi}_P)}{\partial(\theta_1, \theta_2, \ldots, \theta_P)} ,$$

so that we can apply the implicit function theorem to (7). Let $\alpha \triangleq (\theta_1, \theta_2, \ldots, \theta_P)$ and $\beta \triangleq (\theta_{P+1}, \theta_{P+2} \cdots, \theta_{T(T+1)/2})$, so that $\theta = (\alpha, \beta)$ and $\mathring{\theta} = (\mathring{\alpha}, \mathring{\beta})$. Then there is an open neighborhood $\Omega_\alpha \times \Omega_\beta$ of $\mathring{\theta}$ and a unique continuously differentiable function $f: \Omega_\alpha \to \Omega_\beta$ such that $f(\mathring{\beta}) = \alpha$ and

$$\bar{\Phi}_i(f(\beta), \beta) = \bar{\Sigma}_i; \qquad i=1,2,\ldots,P \tag{8}$$

for all $\beta \in \Omega_\beta$. Let us define a new function

$$\Psi_i(\beta) = \bar{\Phi}_i(f(\beta), \beta); \qquad i=1,2,\ldots,T(T+1)/2 \tag{9}$$

Equation (8) implies that $\Psi_1, \Psi_2, \ldots, \Psi_P$ are constants. We will show that $\Psi_{P+1}, \Psi_{P+2}, \ldots, \Psi_{T(T+1)/2}$ are also constant. For $i>P, j>P$, we have

$$\frac{\partial \Psi_i}{\partial \theta_j} = \frac{\partial \bar{\Phi}_i}{\partial \theta_j} + \sum_{k=1}^{P} \frac{\partial \bar{\Phi}_i}{\partial \theta_k} \frac{\partial f_k}{\partial \theta_j} \tag{10}$$

Let us define a neighborhood $\Omega_2=\Omega_1\cap\Omega_\alpha\times\Omega_\beta$ so that rank Φ $=p$ for all $\theta\in\Omega_2$. For each $\theta\in\Omega_1$, each row of $J(\theta)$ is a linear combination of the first p rows so that, assuming the matrix of coefficients $[\lambda_{iq}]$,

$$\frac{\partial\Psi_i}{\partial\theta_j} = \sum_{q=1}^{P} \lambda_{iq}\left\{\frac{\partial\Phi_q}{\partial\theta_j} + \sum_{k=1}^{P} \frac{\partial\Phi_q}{\partial\theta_k}\frac{\partial f_k}{\partial\theta_j}\right\} \tag{11}$$

$$= \sum_{q=1}^{P} \lambda_{iq}\frac{\partial\Psi_q}{\partial\theta_j} = 0,$$

since Ψ_q is constant for $q\leq P$. By (11), Ψ_i is constant for $i=1,2,\ldots,T(T+1)/2$. Thus Φ is constant on the manifold $\{\alpha,\sigma|\alpha=f(\sigma)\}\cap\Omega_2$. Therefore Φ cannot be one-to-one for any neighborhood of the point θ, and θ is not locally identifiable (T). Q.e.d.

Lemma 1, Theorems 1A and 1B combine to form the following, cited in Section 3.6.

Theorem 1 Let $\theta^0\in\mathcal{X}$. Then θ^0 is locally identifiable (T) if and only if, as $n\to\infty$, $[L_{ij}]_{\theta^0}\xrightarrow{a.s.} M>0$.

Genericity

We will show that the regularity condition of Theorem 1 is normally true in the sense that $\mathcal{X}^C$ has Lebesgue measure zero. The following theorem will be needed.

Lemma 2 Let $F(X_1,\ldots,X_n)$ be an analytic function which is not identically zero. The zeros of F are a set of Lebesgue measure zero in the space of $X_1,\ldots,X_n$.

Proof See Fisher [60, pp.163-167].

Theorem 2 The set $\mathcal{R}^C$ has Lebesgue measure zero.

Proof The set $\mathcal{R}^C$ is a subset of $\{\theta | \text{rank } < m\}$. For suppose rank $\underset{\sim}{\Sigma} = m$ at $\hat{\theta}$, so that there is a square submatrix J' of J such that J'>0. By continuity of the elements of J, J'>0 and hence rank $\underset{\sim}{\Sigma} = m$ for all θ is some neighborhood of $\hat{\theta}$. Thus $\hat{\theta}$ is regular. But by Lemma 2 $\{\theta | \text{rank} \underset{\sim}{\Sigma} < m\}$ is of Lebesgue measure zero, since J' is analytic and not identically zero. Q.e.d.

Appendix 8 Listing of Computer Program GCURVE

The following appendix consists of a listing for the approximately 1800 card FORTRAN program written to implement the methodology of this monograph. The program was written specifically for the CDC 6400, Smithsonian Astrophysical Observatory, Cambridge, Mass.

Bibliography

(Not all publications are cited in text)

[1] Bryk, Anthony S. and Weisberg, Herbert I. "Value Added Analysis: A Dynamic Approach to the Estimation of Treatment Effects." Journal of Educational Statistics. Summer 1976. Volume 1, Number 2, pp. 127-155.

[2] Bryk, Anthony S., and Weisberg, Herbert I. Use of Non-Equivalent Control Groups When Subjects are Growing. Huron Institute, Cambridge, Mass., March, 1976. Mimeograph. To appear in Psychological Bulletin, 1977.

[3] Box, George E.P., and Jenkins, Gwilym M. Time Series Analysis. San Francisco: Holden-Day (1970).

[4] Cronbach, L.J., and Furby L. "How Should We Measure 'Change' or Should We?" Psychological Bulletin, 1970, 74, 68-80.

[5] Campbell, D.T., and Stanley, J.C. Experimental and Quasi-experimental Designs for Research. Chicago: Rand-McNally (1963).

[6] Cronbach, L.J., Rogosa, D., Floden, R. and Price, G. Analysis of Covariance: Angel of Salvation or Temptress and Deluder? Stanford Evaluation Consortium. Stanford University, 1976. Mimeograph.

[7] McCall, Robert B., Appelbaum, Jack I., Hogarty, Pamela, S. Developmental Changes in Mental Performance. Monographs of the Society for Research in Child Development. University of Chicago Press, 1973, 38 (3, Serial No. 150).

[8] Bloom, Benjamin S. Stability and Change in Human Characteristics. New York: Wiley (1964)

[9] Ljung, L. "On consistency for prediction error identification methods," Division of Automatic Control, Lund Inst. Tech., Lund, Sweden, Rep. 7405, March 1974.

[10] Ljung, L. "On consistency of prediction error identification methods" in System Identification: Advances and CAse Studies, D.G. Lainiotis and R.K. R.K. Mehrad, Ed. New York: Academic Press (1977).

[11] Ljung, L. "On consistency and identifiability," presented at the Conf. on Stochastic Systems, Lexington, Kentucky, June 1975.

[12] Caines, P.E. "Prediction Error Identification Methods for Stationary Stochastic Processes." IEEE Transactions on Automatic Control, AC-21, August, 1976, pp.500-505.

[13] Åström, Karl J. Introduction to Stochastic Control Theory. Academic Press, New York and London 1970.

[14] Kenny, David A. "A Quasi-Experimental Approach Assessing Treatment Effects in the Non-Equivalent Control Group Design", Psychological Bulletin, 1975, 82, 3, 345-362.

[15] Crano, W.B., Kenny,D.A., and Campbell, D.T. "Does intelligence cause achievement? A cross-logged panel analysis." Journal of Educational Psychology, 1972, 63, 258-275.

[16] Campbell, D.T. "Temporal changes in treatment effect correlations. A quasi-experimental model for institutional records and longitudinal studies." In G.V. Glass (Ed.), Proceedings of the 1970 Invitational Conference on Testing Problems Princeton, N.J.: Educational Testing Service, 1971.

[17] Rosenberg, Barr. "A Survey of Stochastic Parameter Regression," Annals of Economic and Social Measurement, 2/4 1973.

[18] Zellner, A. "On the Aggregation Problem: A New Approach to a Troublesome Problem" Rep. #6628, Center for Math Studies in Business and Economics, Univer. of Chicago. In Economic Models, Estimation and Risk Programming: Essays in Honor of Gerhard Tintner (K.A. Fox et. al., eds.) Springer-Verlag New York, 1969.

[19] Swamy, P.A.V.B. "Linear Models with Random Coefficents" Academic Press, New York and London.

[20] Joreskog, K.G. Factoring the multitest-multioccasion correlation matrix. Research Bulletin 69-62. Princeton, N.J.: Educational Testing Service, 1969.

[21] Joreskog, K.G. A general method for estimating a linear structural equation system. In A.S. Goldgerger, & O.D. Duncan (eds.) Structured equation models in the social sciences. New York: Seminar Press, 1973.

[22] Joreskog, K.G., & van Thillo, M. "LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables." Research Bulletin 72-56. Princeton: Educational Testing Service (December 1972).

[23] Kenny, David A. "Cross-lagged and Synchronous Common Factors in Panel Data". In A.S. Goldberger & O.D. Duncan (eds.), Structured Equation Models in the Social Sciences. New York, Seminar Press, 1973.

[24] Doob, S.L. Stochastic Process New York: Wiley (1953).

[25] Wiener, N. and Mansani, P. "The Prediction Theory of Multivariate Stochastic Processes". Part I. Acta Math., 98 (1957), 111-150. Part II Acta Math., 99 (1958), 93-137.

[26] Mehra, Raman K. "On the Identification of Variances and Adaptive Kalman Filtering". IEEE Transactions on Automatic Control. Vol. AC-15, No.2, April 1970. pp.175-184.

[27] Mehra, Raman K. "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation." AIAA Journal Vol.9, No.1, Jan. 1971. pp.28-31.

[28] Mehra, Raman K. "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering." IEEE Transactions on Automatic Control. Vol. AC-16, No.1, Feb. 1971. pp.12-21.

[29] Kalman, R.E. "A New Approach to Linear Filtering and Prediction Problems" Trans. ASME, Vol.82D, 1960.

[30] Kalman, R.E., and Bucy, R. "New Results in Linear Filtering and Prediction". Trans. ASME, Vol.83D, 1961.

[31] Kendall, M.G., and Stuart, A. The Advanced Theory of Statistics, Vol. 1-3. London: Charles Griffin (1961).

[32] Akaike, H. "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion." In System Identification: Advances and Case Studies. Mehra, R.K. and Lainiotis, D.G. (Eds.) New York: Academic Press (1976).

[33] Lord, F.M., and Novick, M.R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley (1968).

[34] Faurre, P.L. "Stochastic Realization Algorithms" In System Identification: Advance and Case Studies. Mehra, R.K. and Lainiotis, D.G. (Eds.) New York: Academic Press (1976).

[35] Caines, P.E., and Rissanen, J. "Maximum Likelihood Estimation of Parameters in Multivariate Gaussian Stochastic Processes" IEEE Transactions on Information Theory. Vol. IT-20, No. 1, January 1974, pp.102-104.

[36] Wald, A. "Note on the consistency of the maximum likelihood estimate" Ann. Math Statist. Vol 20., pp.595-601, 1949.

[37] Wold, H. A Study in the Analysis of Stationary Time Series. Second Edition. Almquist and Wiksell, Stockholm, 1954.

[38] Harville, David A. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." J.A.S.A. Vol.72, No.358, pp.320-340, June 1977.

[39] Bock, R. Darrel, and Thessen, David M. "Fitting Multi-Component Models for Growth in Stature", Proc. 9th Int. Biometric Conference, Boston, August 1976.

[40] Bock, R.D. Basic issues in the measurement of change. Proceedings of the 2nd International Symposium on Educational Measurement, 1975 (in press).

[41] Bock, R.D., Wainer, H., Petersen, A., Murray, J., Thissen, D. and Roche, A.F. A parameterization for individual human growth curves. Human Biology, 1973, 45, 63-80.

[42] Bock, R.D. and Petersen, A. A multivariate correction for attenuation. Biometrika, 1975, 62, 673-678.

[43] Cootner, Paul (ed.) The Random Character of Stock Market Prices Cambridge: MIT Press (1964).

[44] Sontog, L.W., Baker, C.T., and Nelson, V.L Mental growth and personality development: a longitudinal study. Monographs of the Society for Research in Child Development. 1958, 23 (2, Serial No. 6).

[45] Rissanen, J. "Minimax Entropy Estimation of Models for Vector Processes" In System Identification: Advances and Case Studies. D.G. Lainiotis and R.K. Mehra (eds.). New York: Academic Press (1977).

[46] Layzer, David. "Heritability Analyses of IQ Scores: Science or Numerology?" Science, March 1974, Vol.183, pp.1259-1266.

[47] Campbell, Donald T. One Perspective on Cronbach Day (October 28, 1976) at Northwestern. Mimeographed paper. Northwestern University, January 1977.

[48] Campbell, D.T., and Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (ed.), Compensatory education: A national debate, Vol. 3, Disadvantaged child. New York: Brunner/Mazel, 1970, 185-210. Reprinted: E.L. Struening, and M Guttentag, (Eds.), Handbook of evaluation research. Beverly Hills, Sage Publications, 1975, pp.597-617.

[49] Preschool Inventory Revised Edition--1970: Handbook Princeton: Educational Testing Service, (1970).

[50] Porter, A.C. The effects of using fallible variables in the analysis of covariance (Doctoral dissertation, University of Wisconsin, 1967). Reprint. Ann Arbor, Mich.: University Microfilms (1968).

[51] Taylor, Angus E. Advanced Calculus Boston: Ginn and Company (1955).

[52] Bryk, Anthony S. An Investigation of the Effectiveness of Alternative Statistical Adjustmant Strategies in the Analysis of Quasi-Experimental Growth Data. Doctoral Thesis. Graduate School of Education, Harvard (1977).

[53] Sage, A.P. and Melsa, J.L. Estimation Theory with Applications to Communications and Control. New York: McGraw-Hill (1971).

[54] Bryson, A.E. and Ho, Y.C. Applied Optimal Control, Waltham, Mass.: Ginn-Blaisdell (1969).

[55] Anderson, T.W. An Introduction to Multivariate Statistical Analysis New York: Wiley (1958).

[56] Akaike, H. "Markovian Representation of Stochastic Processes by Canonical Variables." SIAM J. Control. 13, No.1, January 1975.

[57] Van der Voort and Dorpema (MINIM)

[58] Fletcher, R. Fortran Subroutines for Minimization by Quasi-Newton Methods Atomic Energy Research Establishment, Harwell, Berkshire (1972).

[59] Kailath, T. "An innovations approach to least-squares estimation, Part 1", IEEE-Trans.Auto Control, AC-13 (1968),646-660.

[60] Fisher, F.M. The Identification Problem in Econometrics New York: McGraw-Hill (1966).

[61] Cronbach, L.J., G.C. Gleser, H. Nanda, and N. Rajartnam, The Dependibility of Behavioral Measurements. New York: Wiley (1972).

[62] Casti, J.L. Dynamical Systems and Their Applications. New York: Academic Press (1977).

[63] Rissanen, J. "Recursive Identification of Linear Systems" SIAM J. Control, 9, No.3, August 1971.

[64] Ho, B.L. and Kalman, R.E. "Effective Construction of linear state-variable models from input/output function." Proc.Third Allerton Conference, Urbana, Illinois, 1966, pp.449-459.

[65] Akiake, H., "Markovian Representation of Stochastic Processes by Canonical Variables", SIAM J. Control, 13, No.1, January 1975.

[66] Mehra, R.K. "An Algorithm to Solve Matrix Equations $PH^T=G$ and $P=\Phi P\Phi^T+\Gamma\Gamma^T$," IEE Trans actions on Automatic Control, October, 1970.

[67] K.J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records" in Proc. 1965 IFAC Symp. on the Theory of Self-Adaptive Control Systems, 1966.

[68] D.Q. Mayne, "A canonical model for identification of multivariable linear systems," IEEE Trans. Automat.Contr., AC-17, No.5, pp.728-729, 1972.

[69] Wallace, Andrew. Differential Topology. Reading, Mass.: Benjamin, (1968).

[70] Tse, Edison. Information Matrix and Local Identifiability of Parameters. Paper 20-3 in Proc. J.A.C.C., Columbus, 1973.

[71] Tse, Edison and Anton, John J. "On the Identifiability of Parameters" IEEE Trans. A.C. AC-17, No.5, October 1972.

[72] Milnor, J.W. Topology from the Differentiable Viewpoint. Charlottesville: University Press of Virginia, 1965.

[73] Gohberg, I.C. and Krein, M.G., "On the Factorization of Operators in Hilbert Spaces", Amer. Math. Soc. Trans. 51 (1966).

[74] Goodwin, G.C. and Payne, R.L., Dynamic System Identification: Experimental Design and Data Analysis. New York: Academic Press (1977).

[75] Youla, D.C. "On the Factorization of Rational Matrices", IEEE Trans. Inform. Theory IT-7 (1961), 172-184.

[76] Brockett, R.W. Finite Dimensional Linear Systems. New York: Wiley (1970).

[77] Caines, P.E. "Stochastic System Identification" Unpublished Notes. Cambridge (1978).

[78] Anderson, B.D.O. "An Algebraic Solution to the Spectral Factorization Problem", IEEE Trans. in Auto. Control AC-12, pp.410-414, 1967.

[79] Kalman, R.E. "When is a Linear Control System Optimal?", J. Basic Eng. p.51-60, March 1964.

[80] Faurre, P. "Identification par minimisation d'une representation markovienne de processes aleatoire", Symp. on optimization, Nice, June 1968. Berlin: Springer (1969).

[81] Blomqvist, Nils. "On the Relation between Change and Initial Value" J.A.S.A. 1977, 72, No.360, pp.746-749.

[82] Faurre, P. "Realisations markoviennes de processus stationaires" IRIA, Rapport de Recherche 13, March 1973.

[83] Kalman, R.E. "A New Approach to Linear Filtering and Prediction Theory" J. Basic. Eng., March 1960, pp.35-45.

[84] Bayley, N. "Consistency and variability in the growth of intelligence from birth to eighteen years." Psychol. 75, 165-196 (1949).

[85] Mosteller, F. and Tukey, J.W. Data Analysis and Regression. Reading, Mass.: Addison-Wesley (1977).

[86] Tukey, J.W. "Bias and confidence in not-quite large samples." Abstract in Ann. Math. Statist., 29, 614 (1958).

[87] Caines, P.E. and Ljung, L. "Asymptotic Accuracy and Normality of Prediction Error Estimates", Control Systems Report No. 7602, Dept. E.E., U. of Toronto, Feb. 1976. Presented as "Prediction Error Estimators: Asymptotic Normality and Accuracy", Conference on Decision and Control, Clearwater Beach, Fla., December 1976.

[88] Caines, P.E. "Stationary Linear and Nonlinear System Identification and Predictor Set Completeness". Division of Applied Sciences, Harvard University. March 1978.

[89] Abt Associates Inc. National Day Care Study Reports: First Annual Report (1976), Second Annual Report (1977), Second Annual Annual Research Report (1978), Preliminary Findings (1978). Administration for Children, Youth and Families (ACYF), DHEW, Washington, D.C.